

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 10:20:21 ; Search time 341 Seconds

(without alignments)

4796.353 Million cell updates/sec

Title: M64347

Perfect score: 385

Sequence: 1 GACITTCAGCAAGCTGTA.....GTGCATGCTGCCAGAGGTG 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	100.0	3829	6	ABL68525
2	385	100.0	3829	6	ABL68525 Kidney ca
3	385	100.0	3829	7	ABQ88193
4	385	100.0	4093	8	ADA05418 Human FGF
5	385	100.0	4093	9	ADA02847 Human FGF
6	385	100.0	4093	9	AD572585 Human FGF
7	385	100.0	4093	9	AD572585 Human FGF
8	385	100.0	33352	8	ADA02846 Human FGF
9	385	100.0	33352	9	AD572584 Human FGF
10	383.4	99.6	3582	6	ADA040522 Human fib
11	291	75.6	2052	5	AA506078 Angiotens
12	159.2	41.4	33991	8	ADA02843 Mouse Fgf
13	159.2	41.4	33991	9	AD572581 Mouse Fgf
14	159.2	41.4	33991	9	AD572581 Mouse Fgf
15	126.2	32.8	530	9	ADB56952 Toxicity-
16	60	15.6	60	6	AAK39461 Human spl
17	49.4	12.8	238	4	AAK33820 Murine tr
18	40.8	10.9	21234	6	AAK32039 Human kin
19	40.2	10.4	1266	5	AAC88112 Human FLE
20	40.2	10.4	1831	5	ABV27515 Human pro
21	40.2	10.4	1831	5	ABV21694 Human pro
22	40	10.4	1895	4	AAH18061 Human CDN
23	39.8	10.3	700	4	AAH93026 Human inf

ALIGNMENTS

RESULT 1

ABL68525
ID ABL68525 standard; DNA; 3829 BP.

XX AC ABL68525;

XX AC ABL68525;

XX DT 15-MAY-2002 (first entry)

XX DE Kidney cancer related gene sequence SEQ ID NO:6862.

XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

XX KW cytosstatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

XX KW gene; ds.

XX OS Homo sapiens.

XX PN WO200194629-A2.

XX PD 13-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US010838.

XX PR 05-JUN-2000; 2000US-0209473P.

XX PR 05-JUN-2000; 2000US-0209531P.

XX PR 18-SEP-2000; 2000US-0233133P.

XX PR 18-SEP-2000; 2000US-0233617P.

XX PR 20-SEP-2000; 2000US-0234009P.

XX PR 20-SEP-2000; 2000US-0234034P.

XX PR 20-SEP-2000; 2000US-0234052P.

XX PR 22-SEP-2000; 2000US-0234509P.

XX PR 22-SEP-2000; 2000US-0234567P.

XX PR 25-SEP-2000; 2000US-0234923P.

XX PR 25-SEP-2000; 2000US-0235077P.

XX PR 25-SEP-2000; 2000US-0235082P.

XX PR 25-SEP-2000; 2000US-0235134P.

XX PR 26-SEP-2000; 2000US-0235280P.

XX PR 26-SEP-2000; 2000US-0235638P.

XX PR 27-SEP-2000; 2000US-0235711P.

XX PR 27-SEP-2000; 2000US-0235720P.

XX PR 27-SEP-2000; 2000US-0235840P.

XX PR 27-SEP-2000; 2000US-0235863P.

XX PR 28-SEP-2000; 2000US-0236028P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

XX PR 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 28-SEP-2000; 2000US-0236842P.
 PR 28-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVALON) AVALON PHARM.
 XX
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 PI
 XX
 XX WPI; 2002-198264/24.
 XX
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 XX Claim 1; SEQ ID NO 6862; 44pp; English.
 XX
 XX The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 XX
 XX Sequence 3829 BP; 776 A; 1100 C; 1163 G; 790 T; 0 U; 0 Other;
 SQ

Query Match 100.0%; Score 385; DB 6; Length 3829;
 Best Local Similarity 100.0%; Pred. No. 2.1e-94;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATCTCTTAATGCTGTGTCCTCCAGGCA 60
 DB 3336 GACTTCAAGCAAGCTGGTATTTCATACAAATCTCTTAATGCTGTGTCCTCCAGGCA 3395
 QY 61 GGGAGACGGTTTCCAGGAGGGCCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 120
 DB 3396 GGGAGACGGTTTCCAGGAGGGCCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 3455
 QY 121 ACAAGTTTAT 180
 DB 3456 ACAAGTTTAT 3515
 QY 181 AGACTTAACTTCTTACGCAAGCTCTAGAGTTTATAGCTGTGACGCTACCTTTCA 240
 DB 3516 AGACTTAACTTCTTACGCAAGCTCTAGAGTTTATAGCTGTGACGCTACCTTTCA 3575

QY 241 AAGCTTGGAGGAGCCGTGAATTCAGTTGGTTCTGTTCTGTTACTGTTACTGGCCCTGAG 300
 DB 3576 AAGCTTGGAGGAGCCGTGAATTCAGTTGGTTCTGTTCTGTTACTGTTACTGGCCCTGAG 3635
 QY 301 TCTGGCAGCTGTCCCTTGTCTGCTCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
 DB 3636 TCTGGCAGCTGTCCCTTGTCTGCTCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 3695
 QY 361 GCCCAGTGCATGTGGCCAGAGGTG 385
 DB 3696 GCCCAGTGCATGTGGCCAGAGGTG 3720
 RESULT 2
 ABQ88193
 ID ABQ88193 standard; cDNA; 3629 BP.
 XX
 AC ABQ88193;
 XX
 DT 18-SEP-2002 (first entry)
 XX
 DE Human osteoblast differentiation related cDNA SEQ ID NO 100.
 XX
 KW Human; osteoblast; stem cell differentiation; bone tissue deposition;
 KW osteoporosis; osteopathic; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200250301-A2.
 XX
 PD 27-JUN-2002.
 XX
 PF 18-DEC-2001; 2001WO-US048276.
 XX
 PR 18-DEC-2000; 2000US-0255882P.
 PR 24-APR-2001; 2001US-0255691P.
 XX
 PA (GENE-) GENE LOGIC INC.
 PA (PROC) PROCTER & GAMBLE CO.
 XX
 PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
 PI Mertz L;
 XX
 DR WPI; 2002-557663/59.
 XX
 XX Use of genes and their expression profiles associated with osteoblast
 PT differentiation for screening modulators bone formation, for diagnosing
 PT or treating e.g. osteoporosis, or as markers for the differentiation
 PT process.
 PT
 XX
 PS Claim 1; SEQ ID NO 100; 78pp + Sequence Listing; English.
 XX
 CC The invention relates to genes and their expression profiles are used
 CC for: (a) screening modulators of precursor stem cell differentiation into
 CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
 CC deposition of bone tissue, abnormal rate of osteoblast formation or
 CC osteoporosis; or (c) treating or monitoring treatment of the conditions
 CC cited in (b) or monitoring the progression of bone tissue deposition.
 CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
 CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
 CC induced abnormalities in bone formation or bone loss, conditions that
 CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
 CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
 CC or fibrous dysplasia. The present sequence is that of an osteoblast
 CC differentiation associated cDNA marker of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3829 BP; 776 A; 1100 C; 1163 G; 790 T; 0 U; 0 Other;
 Query Match 100.0%; Score 385; DB 6; Length 3829;
 Best Local Similarity 100.0%; Pred. No. 2.1e-94;

Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCACAGGCA 60
Db 3336 GACTTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCACAGGCA 3395
QY 61 GGGAGACGGTTCACAGGAGGGCGCGCCCTGTGTGTCAGGTTCCGATGTTATAGATGT 120
Db 3396 GGGAGACGGTTCACAGGAGGGCGCGCCCTGTGTGTCAGGTTCCGATGTTATAGATGT 3455
QY 121 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 180
Db 3456 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 3515
QY 181 AGACTTAACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGTGTACCTTTCA 240
Db 3516 AGACTTAACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGTGTACCTTTCA 3575
QY 241 AGCTTGGAGGAGGAGCGGGAATTCAGTTGCTTCTGTACTGTTACTGGGCCCTGAG 300
Db 3576 AGCTTGGAGGAGGAGCGGGAATTCAGTTGCTTCTGTACTGTTACTGGGCCCTGAG 3635
QY 301 TCTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 360
Db 3636 TCTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 3695
QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
Db 3696 GCCCAGTGCATGGTGGCCAGAGGTG 3720

RESULT 3
AAD55418
ID AAD55418 standard; DNA; 3829 BP.
AC AAD55418;
XX 07-AUG-2003 (first entry)
XX Human FGFR-3 DNA #2.
XX Human; antisense; fibroblast growth factor receptor 3; prophylaxis;
KW developmental disorder; hyperproliferative disorder; antisense therapy;
KW FGFR-3; ACH; JTK4; CEK2; cancer; chromosome 4p16.3; gene; ds.
XX Homo sapiens.
XX WC2003023004-A2.
XX 20-MAR-2003.
XX 06-SEP-2002; 2002WO-US028549.
XX 10-SEP-2001; 2001US-00953047.
XX (ISIS-) ISIS PHARM INC.
XX Monia BP, Wyatt JR;
XX WPI; 2003-313244/30.
XX Novel compound targeted to a nucleic acid molecule encoding fibroblast
PT growth factor receptor 3, useful for inhibiting the expression of the
PT receptor and for treating an animal having cancer or developmental
PT disorder.
XX Example 15; Page 89-92; 120pp; English.
XX The invention relates to antisense compounds targeted to a nucleic acid
CC molecule encoding fibroblast growth factor (FGF) receptor 3 (also known
CC as FGFR-3, ACH, JTK4 and CEK2) to inhibit its expression. Antisense
CC compounds of the invention are useful for treating diseases or conditions
CC associated with FGFR-3 such as developmental disorders or

CC hyperproliferative disorders, especially cancer of colorectal, bladder,
CC bone, lung, cervical, breast or skin. They are useful as research
CC reagents, therapeutics, prophylaxis, kits and diagnostics, and as tools
CC in differential and/or combinatorial analyses to elucidate expression
CC patterns of a portion of the genes expressed within cells and tissues.
CC They are also useful in antisense therapy. The present sequence is human
CC FGFR-3 DNA
XX
SQ Sequence 3829 BP; 776 A; 1100 C; 1163 G; 790 T; 0 U; 0 Other;
Query Match 100.0%; Score 385; DB 7; Length 3829;
Best Local Similarity 100.0%; Pred. No. 2.1e-94;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCACAGGCA 60
Db 3336 GACTTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCACAGGCA 3395
QY 61 GGGAGACGGTTCACAGGAGGGCGCGCCCTGTGTGTCAGGTTCCGATGTTATAGATGT 120
Db 3396 GGGAGACGGTTCACAGGAGGGCGCGCCCTGTGTGTCAGGTTCCGATGTTATAGATGT 3455
QY 121 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 180
Db 3456 ACAAGTTTATATATCTATATATATAATTTATGAGTTTTCACAGATGTTATTTGTTGT 3515
QY 181 AGACTTAACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGTGTACCTTTCA 240
Db 3516 AGACTTAACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGAGTGTACCTTTCA 3575
QY 241 AAGCTTGGAGGAGGAGCGGGAATTCAGTTGCTTCTGTACTGTTACTGGGCCCTGAG 300
Db 3576 AAGCTTGGAGGAGGAGCGGGAATTCAGTTGCTTCTGTACTGTTACTGGGCCCTGAG 3635
QY 301 TCTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 360
Db 3636 TCTGGCAGCTGTCCCTTGCCTTCAGGCGCCATGGCTCAGGTTGCTCTTCTTGGG 3695
QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
Db 3696 GCCCAGTGCATGGTGGCCAGAGGTG 3720
RESULT 4
ADA02847
ID ADA02847 standard; cDNA; 4093 BP.
XX ADA02847;
XX 06-NOV-2003 (first entry)
XX Human FGFR3 carcinoma associated cDNA, SEQ ID NO:1365.
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ss.
XX Homo sapiens.
XX WO2003057146-A2.
XX 17-JUL-2003.
XX 26-DEC-2002; 2002WO-US041414.
XX 26-DEC-2001; 2001US-00035832.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX WPI; 2003-587068/55.
XX

```
PT New recombinant nucleic acid encoding carcinoma associated protein,
XX useful for preparing compositions for treating carcinomas.
XX
XX Claim 1; SEQ ID NO 1365; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (Cap) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4093 BP; 799 A; 1198 C; 1268 G; 828 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 385; DB 8; Length 4093;
XX Best Local Similarity 100.0%; Pred. No. 2.2e-94;
XX Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTCCTTAATGCTGTGTGCCAGGCA 60
DB 3600 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTCCTTAATGCTGTGTGCCAGGCA 3659
QY 61 GGGAGACGGTTTCAGGAGGGGGCCGCTGTGTGAGTTCGGATGTTATTAGATGTT 120
DB 3660 GGGAGACGGTTTCAGGAGGGGGCCGCTGTGTGAGTTCGGATGTTATTAGATGTT 3719
QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATAT 180
DB 3720 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATAT 3779
QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 240
DB 3780 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 3839
QY 241 AAGCTTGGAGGAGCGCGTAATTCAGTTCGTTCTGCTACTGTACTTGGGCCCTGAG 300
DB 3840 AAGCTTGGAGGAGCGCGTAATTCAGTTCGTTCTGCTACTGTACTTGGGCCCTGAG 3899
QY 301 TCTGGGAGCTGTCCCTTGTCTGCTGAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
DB 3900 TCTGGGAGCTGTCCCTTGTCTGCTGAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 3959
QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 3960 GCCCAGTGCATGGTGGCCAGAGGTG 3984
XX
XX RESULT 5
XX ID ADB72585 standard; mRNA; 4093 BP.
XX AC ADB72585;
XX DT 04-DEC-2003 (first entry)
XX DE Human FGPR3 mRNA.
XX
```

```
KW human; ss; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma.
XX
XX Homo sapiens.
XX
XX WO2003008593-A2.
XX
XX 30-JAN-2003.
XX
XX 26-DEC-2001; 2001WO-US051291.
XX
XX 02-MAR-2001; 2001US-00798586.
XX
XX 23-OCT-2001; 2001US-00004113.
XX
XX 08-NOV-2001; 2001US-00052482.
XX
XX 30-NOV-2001; 2001US-00997722.
XX
XX 20-DEC-2001; 2001US-00034650.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW, Engelhard BK;
XX WPI; 2003-239337/23.
XX
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
XX Claim 1; SEQ ID NO 413; 2304pp; English.
XX
XX The invention relates to a novel recombinant nucleic acid comprising a
XX nucleotide sequence selected from any of the 660 sequences fully defined
XX in the specification. A polynucleotide of the invention has cytostatic
XX activity, and may have a use in gene therapy, or in a vaccine. The
XX recombinant nucleic acids and polypeptides are useful for treating
XX carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX sarcomas. The present sequence represents a human mRNA of the invention.
XX
XX Sequence 4093 BP; 799 A; 1198 C; 1268 G; 828 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 385; DB 9; Length 4093;
XX Best Local Similarity 100.0%; Pred. No. 2.2e-94;
XX Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTCCTTAATGCTGTGTGCCAGGCA 60
DB 3600 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTCCTTAATGCTGTGTGCCAGGCA 3659
QY 61 GGGAGACGGTTTCAGGAGGGGGCCGCTGTGTGAGTTCGGATGTTATTAGATGTT 120
DB 3660 GGGAGACGGTTTCAGGAGGGGGCCGCTGTGTGAGTTCGGATGTTATTAGATGTT 3719
QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATAT 180
DB 3720 ACAAGTTTATATATCTATATATATATATATATATATATATATATATATATAT 3779
QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 240
DB 3780 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 3839
QY 241 AAGCTTGGAGGAGCGCGTAATTCAGTTCGTTCTGCTACTGTACTTGGGCCCTGAG 300
DB 3840 AAGCTTGGAGGAGCGCGTAATTCAGTTCGTTCTGCTACTGTACTTGGGCCCTGAG 3899
QY 301 TCTGGGAGCTGTCCCTTGTCTGCTGAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
DB 3900 TCTGGGAGCTGTCCCTTGTCTGCTGAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 3959
QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 3960 GCCCAGTGCATGGTGGCCAGAGGTG 3984
XX
XX RESULT 6
XX ADB72585
XX AC ADB72585;
XX DT 04-DEC-2003 (first entry)
XX DE Human FGPR3 mRNA.
XX
```


ID ADC85326 standard; DNA; 4093 BP.
XX ADC85326;
AC
XX
DT 01-JAN-2004 (first entry)
XX
XX Human Fgfr3 genomic sequence.
DE
XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
XX Homo sapiens.
OS
XX WO2003045230-A2.
FN
XX
XX 05-JUN-2003.
XX
XX 02-DEC-2002; 2002WO-US038582.
PR
XX 30-NOV-2001; 2001US-00997722.
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW, Engelhard EK;
PI
XX WPI; 2003-513603/48.
DR
XX
XX New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.
PT
PS Claim 1; SEQ ID NO 112; 983pp; English.
XX
XX The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85315-
CC ADC85514 represent CA genes of the invention.
XX
XX Sequence 4093 BP; 799 A; 1198 C; 1268 G; 828 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 385; DB 9; Length 4093;
Best Local Similarity 100.0%; Pred. No. 2.2e-94;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATTCGCTGTGTCGCCAGGCA 60
DB 3600 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATTCGCTGTGTCGCCAGGCA 3659
QY 61 GGGAGACGGTTTCCAGGGAGGGCCGCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 120
DB 3660 GGGAGACGGTTTCCAGGGAGGGCCGCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 3719
QY 121 ACAAGTTTATATATCTAT 180
DB 3720 ACAAGTTTATATATCTAT 3779
QY 181 AGACTTAAACACTTCTACGCAATGCTCTAGAGTTTATAGCTGACGCTACCTTCA 240
DB 3780 AGACTTAAACACTTCTACGCAATGCTCTAGAGTTTATAGCTGACGCTACCTTCA 3839
QY 241 AAGCTTGGAGGGAAGCGCGTAATTCAGTTGGTTCTGTCTGTCTGTCTGTCTGTCTGTCT 300
DB 3840 AAGCTTGGAGGGAAGCGCGTAATTCAGTTGGTTCTGTCTGTCTGTCTGTCTGTCTGTCT 3899
QY 301 TCTGGGACGCTGTCCTTCTGCTTCCGTCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 360
DB 3900 TCTGGGACGCTGTCCTTCTGCTTCCGTCAGGGCCATGGCTCAGGGTGGTCTCTTCTTGGG 3959
QY 361 GCCCAGTGCATGTGGCCAGAGGTG 385

DB 3960 GCCCAGTGCATGTGGCCAGAGGTG 3984.
RESULT 7
ADA02846
ID ADA02846 standard; DNA; 33352 BP.
XX
XX ADA02846;
AC
XX
DT 06-NOV-2003 (first entry)
XX
XX Human FGFR3 carcinoma associated gene, SEQ ID NO:1364.
DE
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX gene; ds.
XX
XX Homo sapiens.
OS
XX WO2003057146-A2.
FN
XX
XX 17-JUL-2003.
PD
XX
XX 26-DEC-2002; 2002WO-US041414.
PF
XX
XX 26-DEC-2001; 2001US-00035832.
PR
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW;
PI
XX WPI; 2003-587068/55.
DR
XX
XX New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
PT
XX Claim 1; SEQ ID NO 1364; 245pp; English.
XX
XX The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 33352 BP; 5993 A; 9766 C; 10292 G; 6073 T; 0 U; 1228 Other;
SQ
Query Match 100.0%; Score 385; DB 8; Length 33352;
Best Local Similarity 100.0%; Pred. No. 4.7e-94;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATTCGCTGTGTCGCCAGGCA 60
DB 24436 GACTTCAAGCAAGCTGTGATTTTCATACAAATTTCTTAATTCGCTGTGTCGCCAGGCA 24495
QY 61 CGGAGACGGTTTCCAGGAGGGCCGCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 120

Db	24496	GGGAGACGGTTTCCAGGAGGGGGCCGGCCCTGTGTGTCAGGTTCCGATGTATTAGATGTT	24555
Qy	121	ACAAGTTTATATATATCTATATATATAAATTTATTCAGTTTTCACAGATGTATTGTTGT	180
Db	24556	ACAAGTTTATATATATCTATATATATAAATTTATTCAGTTTTCACAGATGTATTGTTGT	24615
Qy	181	AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCTTCA	240
Db	24616	AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCTTCA	24675
Qy	241	AAGCTTGGAGGGAAGCCGTCGAATTCAGTTGGTTCGTTCTGTACTGTTACTGGGCCCTGAG	300
Db	24676	AAGCTTGGAGGGAAGCCGTCGAATTCAGTTGGTTCGTTCTGTACTGTTACTGGGCCCTGAG	24735
Qy	301	TCTGGGCAGTGTCCCTTCGTTGCCCTGCGAGGCCCATGGCTCAGGCTGCTCTCTCTGGG	360
Db	24736	TCTGGGCAGTGTCCCTTCGTTGCCCTGCGAGGCCCATGGCTCAGGCTGCTCTCTCTGGG	24795
Qy	361	GCCCACTGCATGTGGCCAGAGGTG	385
Db	24796	GCCCACTGCATGTGGCCAGAGGTG	24820
RESULT 8			
ADB72584			
ID	ADB72584	standard; DNA; 33352 BP.	
XX	XX		
AC	ADB72584;		
XX	XX		
DT	04-DEC-2003	(first entry)	
XX	XX		
DE	Human FGFR3 gene.		
XX	XX		
KW	human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;		
KX	cancer; neoplasm; adenocarcinoma; sarcoma; gene.		
XX	Homo sapiens.		
OS	OS		
XX	WO2003008583-A2.		
XX	30-JAN-2003.		
PD			
XX	26-DEC-2001; 2001WO-US051291.		
XX	02-MAR-2001; 2001US-00798586.		
PR	23-OCT-2001; 2001US-00004113.		
PR	08-NOV-2001; 2001US-00052482.		
PR	30-NOV-2001; 2001US-00937722.		
PR	20-DEC-2001; 2001US-00034650.		
XX	(SAGR-) SAGRES DISCOVERY.		
PA	Morris DW, Engelhard EK;		
XX	WPI; 2003-239337/23.		
PI			
XX	New recombinant nucleic acid, useful for treating carcinomas, lymphomas,		
XX	cancers, neoplasm, adenocarcinoma, or sarcomas.		
DR	Claim 1; SEQ ID NO 412; 2304pp; English.		
XX			
CC	The invention relates to a novel recombinant nucleic acid comprising a		
CC	nucleotide sequence selected from any of the 660 sequences fully defined		
CC	in the specification. A polynucleotide of the invention has cytostatic		
CC	activity, and may have a use in gene therapy, or in a vaccine. The		
CC	recombinant nucleic acids and polypeptides are useful for treating		
CC	carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and		
CC	sarcomas. The present sequence represents a human gene of the invention.		
XX	Sequence 33352 BP; 5993 A; 9766 C; 10292 G; 6073 T; 0 U; 1228 Other;		
SQ			
Query Match 100.0%; Score 385; DB 9; Length 33352;			
Best Local Similarity 100.0%; Pred. No. 4.7e-94;			

Matches	385;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	GACTTCAAAAGCAAGCTGGTATATTTTCATACAAATTCCTTAATTGCTGTGTGCCAGGCA	60						
Db	24436	GACTTCAAAAGCAAGCTGGTATATTTTCATACAAATTCCTTAATTGCTGTGTGCCAGGCA	24495						
Qy	61	GGGAGACGGTTTCAGGAGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATAGATGTT	120						
Db	24496	GGGAGACGGTTTCAGGAGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATAGATGTT	24555						
Qy	121	ACAAGTTTATATATATCTATATATATAATTTATTGAGTTTTTACAAGATGTTATTTGTTGT	180						
Db	24556	ACAAGTTTATATATATCTATATATAATTTATTGAGTTTTTACAAGATGTTATTTGTTGT	24615						
Qy	181	AGACTTAAACACTTCTTACCCAAATGCTTCTAGAGTTTTTATAGCTGGAGTCTACCTTTCA	240						
Db	24616	AGACTTAAACACTTCTTACCCAAATGCTTCTAGAGTTTTTATAGCTGGAGTCTACCTTTCA	24675						
Qy	241	AGCTTTGGAGGAAGCCGCGAATTCAGTTGTTTCGTTCTGTACTGTTACTGGGCCCTGAG	300						
Db	24676	AGCTTTGGAGGAAGCCGCGAATTCAGTTGTTTCGTTCTGTACTGTTACTGGGCCCTGAG	24735						
Qy	301	TCTGGCAGCTGTCCCTTGCTTGCTTGCCCTGCAGGGCCATGGCTCAGGTTGCTCTCTCTGGG	360						
Db	24736	TCTGGCAGCTGTCCCTTGCTTGCTTGCCCTGCAGGGCCATGGCTCAGGTTGCTCTCTCTGGG	24795						
Qy	361	GCCCACTGCATGGTGCCAGAGGTG	385						
Db	24796	GCCCACTGCATGGTGCCAGAGGTG	24820						
RESULT	9								
ADCS5325									
ID	ADC85325	standard; DNA; 33352 BP.							
XX	AC	ADC85325;							
XX	DT	01-JAN-2004 (first entry)							
XX	XX	Mouse Fgfr3 coding sequence.							
DE	XX	Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;							
XX	XX	secreted; transmembrane; intracellular; ds.							
KW	XX	Mus sp.							
XX	OS	WO2003045230-A2.							
XX	PN	05-JUN-2003.							
XX	PD	02-DEC-2002; 2002WO-US038582.							
PF	PF	30-NOV-2001; 2001US-00997722.							
XX	XX	(SAGR-) SAGRES DISCOVERY.							
PR	PR	Morris DW, Engelhard EK;							
XX	XX	WPI; 2003-513603/48.							
XX	XX	New recombinant nucleic acid comprising a nucleotide sequence of any of							
PT	PT	the carcinoma-associated (CA) genes, useful for screening for drug							
PT	PT	candidates for diagnosing or treating carcinomas.							
XX	XX	Claim 1; SEQ ID NO 111; 983pp; English.							
XX	XX	The invention relates to a recombinant nucleic acid comprising a							
CC	CC	nucleotide sequence selected from any of the fully defined carcinoma-							
CC	CC	associated (CA) genes from the 50 tables given in the specification. The							
CC	CC	CA proteins are secreted, transmembrane or intracellular proteins. The							
CC	CC	recombinant nucleic acids are useful for screening for drug candidates							
CC	CC	for diagnosing or treating carcinomas. Sequences given in ADC85215-							
CC	CC	ADC85514 represent CA genes of the invention.							

WPI: 2002-519881/55.
P-PSDB; AAR24993.

Stimulating a population of stem cells to differentiate into osteoblast cells useful for treating osteoporosis, by contacting the cells with agent which increases fibroblast growth receptor 3 expression or activity.

Disclosure; Page 49-52; 58pp; English.

The invention relates to a method for stimulating a population of stem cells to differentiate into osteoblast cells by contacting the population with an agent which increases fibroblast growth factor receptor 3 (FGFR3) expression or activity, where increase in FGFR3 protein expression or activity results in differentiation of the stem cells into osteoblast cells. The method is useful for stimulating the population of stem cells to differentiate into osteoblast cells. The method is useful for increasing bone density. The method is useful for screening the agent that modulates the differentiation of population into osteoblast cells, increases bone density, or ameliorates the effects of osteoporosis. The method is useful for diagnosing a condition characterised by abnormal stem cell differentiation, bone density or rate of osteoblast formation and treating a patient with a condition characterised by an abnormal rate of osteoblast formation, bone density or osteoporosis. The present sequence is human FGFR3 DNA

Sequence 3582 BP; 705 A; 1032 C; 1105 G; 740 T; 0 U; 0 Other;

Query Match 99.6%; Score 383.4; DB 6; Length 3582;
Best Local Similarity 99.7%; Pred. No. 5.7e-94;
Matches 384; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAAATGCTGTGTGCCAGGCA 60
3119 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAAATGCTGTGTGCCAGGCA 3178

61 GCGAGACGGTTTCCAGGAGGGCGCGCCCTGTGCGAGGTTCGGATGTTATTAGATGTT 120
3179 GCGAGACGGTTTCCAGGAGGGCGCGCCCTGTGCGAGGTTCGGATGTTATTAGATGTT 3238

121 ACAAGTTTATATATATCTATATATATAAATTTATTGAGTTTTTACAAGATGATTTGTGT 180
3239 ACAAGTTTATATATATCTATATATAAATTTATTGAGTTTTTACAAGATGATTTGTGT 3298

181 AGACTTAACTTCTTACGCAATGCTCTAGAGTTTATAGCTGGAGCTGTACCTTTCA 240
3299 AGACTTAACTTCTTACGCAATGCTCTAGAGTTTATAGCTGGAGCTGTACCTTTCA 3358

241 AAGCTTGGAGGGAAGCGTGCAATTCAGTTGGTTCGTTCTGTACTGTACTGGGCCCTGAG 300
3359 AAGCTTGGAGGGAAGCGTGCAATTCAGTTGGTTCGTTCTGTACTGTACTGGGCCCTGAG 3418

301 TCTGGGCAGTGTGCCCTTGTTCGCTGCAGGGCCATGGCTCAGGGTGCGTCTCTCTGGG 360
3419 TCTGGGCAGTGTGCCCTTGTTCGCTGCAGGGCCATGGCTCAGGGTGCGTCTCTCTGGG 3478

361 GCCCAGTGCATGGTGCCACAGGTG 385
3479 GCCCAGTGCATGGTGCCACAGGTG 3503

RESULT 11
ID AAS06078 standard; DNA; 2052 BP.
XX AAS06078;
AC AC
XX XX
DT 12-SEP-2001 (first entry)
XX XX
DE Angiotensin converting enzyme (ACEV) splice variant DNA #78.
XX Angiotensin converting enzyme splice variant; ACEV; interleukin 6;
KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;

KW platelet-derived endothelial cell growth factor; cardiovascular disease;
 KW cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C; ds;
 KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;
 KW myocardial infarction; coronary arterial thrombosis; renal disease;
 KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;
 KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;
 KW nonaroidotic pulmonary granulomatous disease; endothelial abnormality;
 KW vascular disorder; asbestosis.
 XX
 OS Homo sapiens.
 XX
 PN WO200136632-A2.
 XX
 PD 25-MAY-2001.
 XX
 PF 17-NOV-2000; 2000WO-IL000766.
 XX
 PR 17-NOV-1999; 99IL-00132978.
 PR 10-DEC-1999; 99IL-00133455.
 XX
 XX (COMP-) COMPUGEN LTD.
 XX
 PA Levine Z, David A, Azar I, Khosravi R, Bernstein J;
 PI
 XX
 XX WPI; 2001-336004/35.
 DR P-PSDB; AAU02978.
 XX
 XX Novel alternative splicing variants e.g. variant of angiotensin
 PT converting enzyme (ACEV), useful in identifying candidate compounds
 PT capable of binding to the variant and to detect anti-variant antibodies.
 XX
 PS Claim 1; Page 368-369; 519pp; English.
 XX
 CC The sequence represents a DNA encoding an angiotensin converting enzyme
 CC splice variant (ACEV) polypeptide. The polypeptides of the invention
 CC include variants of granulocyte colony stimulating factor receptor,
 CC glucagon, interleukin 6, platelet-derived endothelial cell growth factor,
 CC cyclin-dependent kinase inhibitor 1C, cellular tumour antigen P53, and
 CC vasoactive intestinal polypeptide receptor 2. The polypeptides and their
 CC associated nucleic acids are useful for identification of variant
 CC sequences and detection of candidate compounds capable of binding the
 CC molecules. The sequences of the invention can be used in the treatment
 CC and diagnosis of various disorders including cardiovascular diseases such
 CC as arteriosclerosis, myocardial infarction and coronary arterial
 CC thrombosis, renal diseases such as diabetic nephropathy, muscular
 CC diseases such as hypertrophy, immune disorders such as immune complex
 CC nephritis, multiple sclerosis, cancer, sarcoidosis, nonaroidotic
 CC pulmonary granulomatous diseases such as asbestosis and vascular
 CC pathologies involving an endothelial abnormality such as deep vein
 CC thrombosis.
 XX
 SQ Sequence 2052 BP; 379 A; 619 C; 667 G; 385 T; 0 U; 2 Other;
 Query Match 75.6%; Score 291; DB 5; Length 2052;
 Best Local Similarity 100.0%; Pred. No. 6.2e-69;
 Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 95 TGCAGGTTCCGATGTTATGAGTTACAGTTTATATATATATATATATATATATAT 154
 DB 1681 TGCAGGTTCCGATGTTATGAGTTTACAGTTTATATATATATATATATATATAT 1740
 QY 155 GAGTTTTTCAAGATGTTATTTGTTAGACTTAACTTCTTACGCAATGCTTCTAGAGT 214
 DB 1741 GAGTTTTTCAAGATGTTATTTGTTAGACTTAACTTCTTACGCAATGCTTCTAGAGT 1800
 QY 215 TTATAGCTGGAGTCTGCTACTTTTCAAAGCTTGGAGGGAAGCGTGAATTCAGTTGGTTC 274
 DB 1801 TTATAGCTGGAGTCTGCTACTTTTCAAAGCTTGGAGGGAAGCGTGAATTCAGTTGGTTC 1860
 QY 275 GTTCTGTACTGTGTACTGGGGCCCTGAGTCTGGGCGAGCTGTCCCTTGTCTCCCTGCGAGGGCC 334
 DB 1861 GTTCTGTACTGTGTACTGGGGCCCTGAGTCTGGGCGAGCTGTCCCTTGTCTCCCTGCGAGGGCC 1920

QY 335 ATGGCTCAGGTTGCTCTCTTCTTGGGGCCCACTGCTGCTGCGCAGAGGTG 385
 DB 1921 ATGGCTCAGGTTGCTCTCTTCTTGGGGCCCACTGCTGCTGCGCAGAGGTG 1971
 RESULT 12
 ADA02843
 ID ADA02843 standard; DNA; 33991 BP.
 XX
 AC ADA02843;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Mouse Egr3 carcinoma associated gene, SEQ ID NO:1361.
 XX
 KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
 KW Gene; ds.
 XX
 OS Mus sp.
 XX
 PN WO2003057146-A2.
 XX
 PD 17-JUL-2003.
 XX
 PF 26-DEC-2002; 2002WO-US041414.
 XX
 PR 26-DEC-2001; 2001US-00035832.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW;
 XX
 XX WPI; 2003-587068/55.
 DR
 XX
 PT New recombinant nucleic acid encoding carcinoma associated protein,
 PT useful for preparing compositions for treating carcinomas.
 XX
 PS Claim 1; SEQ ID NO 1361; 245pp; English.
 XX
 CC The invention relates to recombinant carcinoma associated (CA) nucleic
 CC acid sequences from mouse and human (ADA01462-ADA03094), and to
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The
 CC invention also encompasses expression vectors and host cells comprising a
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
 CC binds to the protein, and a biochip comprising CA nucleic acid or
 CC fragments thereof. The sequences of the invention were identified using
 CC oncogenic retroviruses, which insert into the genome of the host organism
 CC at random. Many of these do not carry transduced host oncogenes or
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
 CC direct consequence of the effects of proviral integration into host
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
 CC leukaemia) or a propensity to carcinoma by determination of the sequence
 CC of a CA gene, or by determination of CA gene expression in particular
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as
 CC therapeutic agents and in screening and evaluating drug candidates. The
 CC present sequence represents a specifically claimed murine CA nucleic acid
 CC sequence of the invention. Note: The complete sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 33991 BP; 7578 A; 8539 C; 9524 G; 8350 T; 0 U; 0 Other;

Query Match 41.4%; Score 159.2; DB 8; Length 33991;
 Best Local Similarity 78.8%; Pred. No. 1.2e-32;
 Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATTTCTTCTTAATTTGCTGTGTGCCAGGAGGAGGTTTCCAGGGAG 80
 DB 24786 TTTTCATACAGATTTATTTCTTAATTTGCTGTGTGCCAGGAGGAGGTTTCCAGGGAG 24844

QY 81 GGGCGGCGCTGTGCGAGGTTCCGATGTTATTAGATGTTACAGTTTATATATATCTAT 140
 Db 24845 GGGCTGGCGCTGCAGAAAGGTTTCAGATGTTAATAGTTAT----CTGTTACAAGTTTATCT 24900
 QY 141 ATATATAATTTATTGAGTTTTCACAAAGATGTTATTTGTTGTAGACTTAAACACTTCTTACGC 200
 Db 24901 ATCTATATTTATTGAGTTTTCACAAAGTTGT-TTGTGCTGAGGCTTAAACACTTCTTACGC 24959
 QY 201 AATGCTTCAGAGTTTATAGCTGAGCTGCTACCTTTCAAGCTTGGAGGAGCCGCTG 260
 Db 24960 AGTGTCTTCAGACTTTTATAGCTGAGCTGCTACCTTTCAAGCTTGGAGGAGCTGCTG 25019
 QY 261 AATTC-AGTTGCTTCGTTCTGTTACTGTTACTGGGCGCTGAGTCTGGGCGCTGAGTCTGGCTGCTGCTG 319
 Db 25020 AATGCAATTTTGTACTTTTGTACTGTTACTGGGCGCTGAGTCTGGGCGCTGAGTCTGGCTGCTGCTG 25079
 QY 320 CTGCGC 325
 Db 25080 CCTGTC 25085

RESULT 13
 ADB72581
 ID ADB72581 standard; DNA; 33991 BP.

XX ADB72581;
 DT 04-DEC-2003 (first entry)
 DE Mouse Fgfr3 gene.
 XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Mus sp.
 XX WO2003008583-A2.

XX 30-JAN-2003

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00397722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 409; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytostatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a mouse gene of the invention.

XX Sequence 33991 BP; 7578 A; 8539 C; 9524 G; 8350 T; 0 U; 0 Other;

Query Match 41.4%; Score 159.2; DB 9; Length 33991;
 Best Local Similarity 78.8%; Pred. No. 1.2e-32;
 Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATTTCTTAAATTTCTGTGTGTCAGGCGAGGAGCGGTTTCCAGGGAG 80
 Db 24786 TTTTCATACAGATTTATTCATTTGCTATGTTATTCAGGAG-GAGCTGTGCCAGGGAA 24844
 QY 81 GGGCGGCGCTGTGCGAGGTTCCGATGTTATTAGATGTTACAAGTTTATATATATCTAT 140
 Db 24845 GGGCTGGCGCTGCAGAAAGGTTTCAGATGTTAATAGTTAT----CTGTTACAAGTTTATCT 24900
 QY 141 ATATATAATTTATTGAGTTTTCACAAAGATGTTATTTGTTGTAGACTTAAACACTTCTTACGC 200
 Db 24901 ATCTATATTTATTGAGTTTTCACAAAGTTGT-TTGTGCTGAGGCTTAAACACTTCTTACGC 24959
 QY 201 AATGCTTCAGAGTTTATAGCTGAGCTGCTACCTTTCAAGCTTGGAGGAGCCGCTG 260
 Db 24960 AGTGTCTTCAGACTTTTATAGCTGAGCTGCTACCTTTCAAGCTTGGAGGAGCTGCTG 25019
 QY 261 AATTC-AGTTGCTTCGTTCTGTTACTGTTACTGGGCGCTGAGTCTGGGCGCTGAGTCTGGCTGCTGCTG 319
 Db 25020 AATGCAATTTTGTACTTTTGTACTGTTACTGGGCGCTGAGTCTGGGCGCTGAGTCTGGCTGCTGCTG 25079
 QY 320 CTGCGC 325
 Db 25080 CCTGTC 25085

RESULT 14
 ADC85322
 ID ADC85322 standard; DNA; 33991 BP.

XX ADC85322;

XX 01-JAN-2004 (first entry)

XX Human Map3k8 coding sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
 KW secreted; transmembrane; intracellular; ds..

XX Homo sapiens.

XX WO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002WO-US038582.

XX 30-NOV-2001; 2001US-00397722.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of
 PT the carcinoma-associated (CA) genes, useful for screening for drug
 PT candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 108; 983pp; English.

XX The invention relates to a recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the fully defined carcinoma-
 CC associated (CA) genes from the 50 tables given in the specification. The
 CC CA proteins are secreted, transmembrane or intracellular proteins. The
 CC recombinant nucleic acids are useful for screening for drug candidates
 CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
 CC ADC85514 represent CA genes of the invention.

XX Sequence 33991 BP; 7578 A; 8540 C; 9523 G; 8350 T; 0 U; 0 Other;

Query Match 41.4%; Score 159.2; DB 9; Length 33991;
 Best Local Similarity 78.8%; Pred. No. 1.2e-32;
 Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATCAAAATCTTCTTAATCTGTGTGTCCTCCAGGAGAGAGAGGTTTCAGGGAG 80
 DB 24786 TTTTCATACAGATTATCTTAATCTGTGTGTATTCAGGAGAGAGGTTTCAGGGAG 24844
 QY 81 GGGCGGGCCCTGTGTGAGGTTCCGATGTTATTAGATGTTTACAAGTTTATATATATATCTAT 140
 DB 24845 GGGCTGGCCCTGCAAGAAGTTTCAGATGTTTAAATAGTTAT---CTGTACAGTTTATCT 24900
 QY 141 ATATATAATTTTATGAGTTTTCACAGATGTTTGTGTAGATTTTAACTTTCCTTACGC 200
 DB 24901 ATCTATAATTTTATGAGTTTTCACAGATGTTTGTGTAGATTTTAACTTTCCTTATGC 24959
 QY 201 AATGCTTCTAGAGTTTATAGCTGGAGTCTGCTACCTTTTCAAGCTTGGAGGAGCCGTG 260
 DB 24960 AGTCTTCTAGACTTTTATAGCTTACCTGCTACCTTTTCAAGCTTGGAGAGAGTGGTG 25019
 QY 261 AATTC-AGTTGGTTCGTCTGTACTGTACTGTGGCCCTGAGTCTGGGCACTGTCCCTTG 319
 DB 25020 AATGCAATTTTGTACTTTTGTACTGTCTGCTGAGTGGCCCTGAGTCTGGGCTGTCCCTTG 25079
 QY 320 CTGGCC 325
 DB 25080 CCTGTC 25085

RESULT 15
 ID ADB56952/c
 XX ADB56952 standard; DNA; 530 BP.
 AC ADB56952;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Toxicity-related gene, SEQ ID 1978.
 XX
 KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;
 KW drug screening; toxicity assay; ds.
 XX
 OS Unidentified.
 XX
 FN WO2003064624-A2.
 XX
 PD 07-AUG-2003.
 XX
 PF 31-JAN-2003; 2003WO-US003194.
 XX
 PR 31-JAN-2002; 2002US-00060087.
 PR 15-MAR-2002; 2002US-0364045P.
 PR 15-MAR-2002; 2002US-0364055P.
 PR 30-DEC-2002; 2002US-0436643P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 XX
 DR WPI; 2003-689530/65.
 XX
 PT Predicting a toxic effect of a compound, useful in identifying toxicity
 PT markers in liver tissues or cells for drug screening and toxicity assays,
 PT comprises preparing gene expression profile of tissue or cells exposed to
 PT the compound.
 XX
 PS Claim 1; SEQ ID NO 1978; 1156pp; English.
 XX
 CC The present invention relates to a method for predicting a toxic effect
 CC of a compound. The method comprises preparing a gene expression profile
 CC of a tissue or cell sample exposed to the compound, and comparing the
 CC gene expression profile to a database comprising SEQ ID 1-4925, where
 CC differential expression of the gene indicates at least one toxic effect.
 CC The method is useful for predicting at least one toxic effect of a
 CC compound, predicting hepatotoxicity or the progression of a toxic effect
 CC of a compound, identifying an agent that modulates the onset or

CC progression of a toxic response, predicting the cellular pathways that a
 CC compound modulates in a cell, and identifying an agent that modulates at
 CC least one activity of a protein. The method and compositions of the
 CC present invention using a database of genes having liver toxin-induced
 CC differential expression, are useful in identifying toxicity markers in
 CC liver tissues or cells for drug screening and toxicity assays. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 530 BP; 173 A; 124 C; 94 G; 139 T; 0 U; 0 Other;

Query Match 32.8%; Score 126.2; DB 9; Length 530;

Best Local Similarity 69.5%; Pred No. 2.5e-24;

Matches 230; Conservative 0; Mismatches 93; Indels 8; Gaps 4;

QY 1 GACTTCAAGCAAGCTGCTA-TTTTCATACAAATCTTCTAATGCTGTGTCCAGGC 59

DB 475 GATTTCAAGGTCAACTGGTA-TTTTCATACAGATTTATCTAATGTTATGTTCCAGGC 416

QY 60 AGGGAGAGCGTTTCCAGGAGGGCCGCTGTGTGCGAGGTCGGATGTTTAGATGT 119

DB 415 AGGGCCCTGTGCCCCAGGAGGGGCTGGCCCTGCAAGAGGTTGAGTTAATAGATGT 356

QY 120 TACAAAGTTTATATATATCTATATATATATAATTTATGAGTTTTCACAGATGTTTGTG 179

DB 355 ---CTGTTACAAGTTTATCTATCTATATAATTTATGAGTTTCTACAAGTTGT-TTTGCTG 301

QY 180 TAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCTTTC 239

DB 300 TAGGCTTAACACTTCTTATGCAAGTATCTAGACTTTTATAGCCTGACTGCTACCTTTC 241

QY 240 AAAGCTTGGAGGAGCGCTGCAATTCA--GTTGGTTCGTTCTGTACTGTTACTGGGCCT 297

DB 240 AAAGCTTGGAGAGAGTCTGCAATGCAATTTTGTGACAGGTTTACTGTTACTGGGCCT 181

QY 298 GAGTCTGGGAGCTGTCCCTTGTGCTGC 328

DB 180 AGGCTTGGTGGCTGTCCCTTGTGCTGC 150

Search completed: September 22, 2004, 11:20:40

Job time : 345 secs

QY	1	GACTTCAAAGCAAGCTGGTATTTTTCATACAAATTTCTTAATTGCTGTGTGCTCCAGGCA	60
DB	24436	GACTTCAAAGCAAGCTGGTATTTTTCATACAAATTTCTTAATTGCTGTGTGCTCCAGGCA	24495
QY	61	GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTGCGAGTTCGATGTTTATTAGATGTT	120
DB	24496	GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTGCGAGTTCGATGTTTATTAGATGTT	24555
QY	121	ACAAGTTTATATATATCTATATATATATAATTTATTTAGTATTTTACAAAGATGTTTGTGT	180
DB	24556	ACAAGTTTATATATATCTATATATATATAATTTATTTAGTATTTTACAAAGATGTTTGTGT	24615
QY	181	AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGTGACTGCTACCTTTCA	240
DB	24616	AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGTGACTGCTACCTTTCA	24675
QY	241	AAGCTTGGAGGGAAGCGTGAATTCAGTGGTTCGTCTGTACTGTTACTGGCCCTGAG	300
DB	24676	AAGCTTGGAGGGAAGCGTGAATTCAGTGGTTCGTCTGTACTGTTACTGGCCCTGAG	24735
QY	301	TCTGGGAGCTGCTCCCTTGTCTGCTGCGAGGCGCATGGCTCAGGTTGGTCTCTTTGGG	360
DB	24736	TCTGGGAGCTGCTCCCTTGTCTGCTGCGAGGCGCATGGCTCAGGTTGGTCTCTTTGGG	24795
QY	361	GCCAGTGCATGGTGGCCAGAGGTG	385
DB	24796	GCCAGTGCATGGTGGCCAGAGGTG	24820
RESULT 11			
US-10-450-859-1			
; Sequence 1, Application US/10450859			
; Publication No. US20040109850A1			
; GENERAL INFORMATION:			
; APPLICANT: Jaiswal, Neelam			
; APPLICANT: Houghton, Adam			
; APPLICANT: Mertz, Lawrence			
; APPLICANT: Ji, Darren			
; APPLICANT: Cook, Jonathan S.			
; APPLICANT: Axelrod, Douglas W.			
; TITLE OF INVENTION: Treatment of Bone Disorders by Modulation of FGFR3			
; FILE REFERENCE: 44921-5078-WO			
; CURRENT APPLICATION NUMBER: US/10/450,859			
; CURRENT FILING DATE: 2003-06-18			
; PRIOR APPLICATION NUMBER: US 60/255,882			
; PRIOR FILING DATE: 2000-12-18			
; PRIOR APPLICATION NUMBER: US 60/285,691			
; PRIOR FILING DATE: 2001-04-24			
; PRIOR APPLICATION NUMBER: US 60/306,879			
; PRIOR FILING DATE: 2001-07-23			
; PRIOR APPLICATION NUMBER: US 60/317,974			
; PRIOR FILING DATE: 2001-09-10			
; NUMBER OF SEQ ID NOS: 14			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 3592			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (99)..(1679)			
; OTHER INFORMATION: FGFR3, GenBank Accession No. XM_017699			
US-10-450-859-1			
Query Match 99.6%; Score 383.4; DB 17; Length 3592;			
Best Local Similarity 99.7%; Pred. No. 6.8e-93;			
Matches 384; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	GACTTCAAAGCAAGCTGGTATTTTTCATACAAATTTCTTAATTGCTGTGTGCTCCAGGCA	60
DB	3119	GACTTCAAAGCAAGCTGGTATTTTTCATACAAATTTCTTAATTGCTGTGTGCTCCAGGCA	3178
QY	61	GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTGCGAGTTCGATGTTTATTAGATGTT	120

; ORGANISM: Mus musculus
US-09-997-722-109

Query Match 41.4%; Score 159.2; DB 12; Length 33991;
Best Local Similarity 78.8%; Pred. No. 9e-32;
Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;
QY 21 TTTTCATACAAATCTCTTAATTGCTGTGTCCTCCAGGCGAGGAGACGGTTTCCAGGGAG 80
DB TTTTCATACAAATCTCTTAATTGCTGTGTCCTCCAGGCGAGGAGACGGTTTCCAGGGAG 24844
QY 81 GGGCCGGCCCTGTGTCAGGTTCCGATGTTATTAGATGTTTACAAGTTTATATATATATCTAT 140
DB GGGCTGGCCCTGCAAGAAGTTTCAGATGTTAATAGTTAT---CTGTTACAAGTTTATCT 24900
QY 141 ATATATAATTTATTGAGTTTTCACAGATGTTTCTGTAGACTTACACATCTCTTACGC 200
DB ATCTATAATTTATTGAGTTTTCACAGATGTTTCTGTAGACTTACACATCTCTTACGC 24959
QY 201 AATGCTTCTAGAGTTTATAGCTTGACGCTGACCTTTCAAAGCTTTGGAGGAAGCCGTG 260
DB AGTGCTTCTAGACTTTTATAGCTGACGCTGACCTTTCAAAGCTTTGGAGACAGTGGTG 25019
QY 261 AATTC-AGTGTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 319
DB AATGCAATTTGTTACTTTGTTACTGCTCACTGGGCGCTAGGCTGGGTGGCTGCTGCTGCTG 25079
QY 320 CTTGCC 325
DB 25080 CTTGTC 25085

RESULT 15
US-10-085-783A-51002
; Sequence 51002, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 51002
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-51002

Query Match 40.2%; Score 154.8; DB 13; Length 239;
Best Local Similarity 98.7%; Pred. No. 9.8e-32;
Matches 156; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 220 AGCGTGACTGTACCTTTCAAAGCTTGAGGGAAGCCGTAATTCAGTTGGTTGTTCT 279
DB 1 AGCGTGACTGTACCTTTCAAAGCTTGAGGGAAGCCGTAATTCAGTTGGTTGTTCT 60
QY 280 GTACTGTTACTGGGCGCTGAGTCTGGGCGAGCTGTCCCTTGTCTGCTGCTGAGGGCCATGGC 339
DB 61 GTACTGTTACTGGGCGCTGAGTCTGGGCGAGCTGTCCCTTGTCTGCTGCTGAGGGCCATGGC 120
QY 340 TCAGGGTGTCTCTTCTTGGGCGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 377
DB 121 TCAGGGTGTCTCTTCTTGGGCGCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 158

Search completed: September 22, 2004, 12:39:28
Job time : 357 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 10:20:21 ; Search time 1784 Seconds
(without alignments)
9353.738 Million cell updates/sec

Title: M64347
Perfect score: 385
Sequence: 1 GACTTCAAAGCAAGCTGTA.....GTGCATGGTGGCCAGAGGTG 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rdt: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	385	100.0	3829	6	AX336353	AX336353 Sequence
2	385	100.0	3829	9	HUMFGFLR	M64347 Human novel
3	385	100.0	4093	6	AX695738	AX695738 Sequence
4	385	100.0	5359	9	BSM805784	BSM805784 Homo sapi
5	385	100.0	16976	9	AF487554	AF487554 Homo sapi
6	385	100.0	33352	6	AX695737	AX695737 Sequence
7	385	100.0	189223	9	AC016773	AC016773 Homo sapi
8	291	75.6	2052	6	AX147524	AX147524 Sequence
9	189	49.1	67890	2	AC099840	AC099840 Homo sapi
10	159.2	41.4	33991	6	AX695734	AX695734 Sequence
11	159.2	41.4	201973	2	AC079504	AC079504 Mus muscu
12	159.2	41.4	245130	2	AC079571	AC079571 Mus muscu
13	157.6	40.9	1848	10	MUSEGFR17	L42132 Mus muscu
14	157.6	40.9	4156	10	BC053056	BC053056 Mus muscu
15	157.6	40.9	4158	10	MUSMER3	M81342 BALB/c fibr
16	141	36.6	180035	2	AC125819	AC125819 Rattus no
17	141	36.6	209428	2	AC125948	AC125948 Rattus no
18	131	34.0	222009	2	AC133613	AC133613 Rattus no
19	131	34.0	267692	2	AC135138	AC135138 Rattus no
20	126.2	32.8	175164	2	AC119911	AC119911 Mus muscu
21	74.4	19.3	299974	2	AC079512	AC079512 Mus muscu
22	68.8	17.9	156848	9	AC097710	AC097710 Homo sapi
23	68.8	17.9	160918	2	AC069196	AC069196 Homo sapi
24	49.4	12.8	238	6	AX524355	AX524355 Sequence
25	49.4	12.8	238	6	AX553093	AX553093 Sequence
26	44	11.4	110000	2	PFMAL13_00	AL844509 Plasmodiu
27	44	11.4	110000	2	PFMAL13_01	Continuation (2 of
28	44	11.4	192581	2	PFMAL13P1	AL049180 Plasmodiu
29	43.6	11.3	148926	2	AC102641	AC102641 Mus muscu
30	43.4	11.3	250195	3	BE014831	BE014831 Plasmodiu
31	43.2	11.2	144022	5	EX088594	EX088594 Zebrafish
32	43	11.2	257426	2	EX546500	EX546500 Danio rer
33	42.8	11.1	184798	2	AC139981	AC139981 Rattus no
34	42.8	11.1	234967	2	AC134282	AC134282 Rattus no
35	42.6	11.1	248294	2	AC097690	AC097690 Rattus no
36	42.6	11.1	120465	2	AC137001	AC137001 Oryza sat
37	42.6	11.1	183046	8	AC087599	AC087599 Oryza sat
38	42.6	11.1	286362	2	AC141488	AC141488 Rattus no
39	42.6	11.1	304969	8	AE017120	AE017120 Oryza sat
40	42.4	11.0	110000	2	AC119727_1	Continuation (2 of
41	42.4	11.0	110000	2	AC119727_2	Continuation (3 of
42	42.4	11.0	110000	2	PFMAL7P1_09	Continuation (10 o
43	42.4	11.0	110000	2	PFMAL7P1_10	Continuation (11 o
44	42.4	11.0	168963	2	AC09705	AC09705 Homo sapi
45	42.4	11.0	169102	5	AL929146	AL929146 Zebrafish

ALIGNMENTS

RESULT 1
AX336353
LOCUS AX336353 3829 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 6862 from Patent WO0194629.
ACCESSION AX336353
VERSION AX336353.1 GI:18127072
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature


```

join(<5018, .5287,5511, .5576,7131, .7300,7384, .7507,
7599, .7789,9456, .9600,10094, .10284,10588, .10733,
11119, .11240,11323, .11433,11514, .11704,11815, .11937,
12021, .12091,12310, .12447,12593, .12698,12880, .>13026)
/gene="FGFR3"
/product="fibroblast growth factor receptor 3"
/notes="alternatively spliced"
join(<5018, .5287,5511, .5576,7131, .7300,7384, .7507,
7599, .7789,9678, .8828,10094, .10284,10588, .10733,
11119, .11240,11323, .11433,11514, .11704,11815, .11937,
12021, .12091,12310, .12447,12593, .12698,12880, .13026)
/gene="FGFR3"
/notes="alternatively spliced"
/codon_start=3
/product="fibroblast growth factor receptor 3"
/protein_id="AAM22079.1"
/db_xref="GI:20452381"
/translation="VPGPPGQEOQLVFGSDAVELSCPPPGGPGMGPTVWKDGTGL
VPSERVLPQRLQVLNASHESGAYSCQRQLTORVLCHFSVRVTDAPSSGDDEDED
EAEQTDVDTGAPYWRPERMDKLLAVPAANTVRFCPAAGNTRQITLTDVLERSPHRIL
BHGRIKLRHQMSLVMSVSPSDEGRNCTCVENKFGSIRQTYTLDVLERSPHRIL
QAGPANTQAVLSEVFEHCYVSDAQPHIQMLKHEVNGSKVGPDTPVTYVLKSMI
SISVADILYGLVARNSERDGELVCRAINFVGAEKAFMSVGHSPRAAEELVEADEA
GESVAGILSYGVGFLLVIAAVALTCLRLSPPKGLGSPTVHKISRFPLKROVLES
NASKSNTPPLVARIARLSSGEGPTLVANSELELPADPKWELSRALTGLKPGHKTINLLG
VMAEALGIDKDRAPQVTVAVKMLKDDATDKDLDLSEMEMMKMILGHKNTINLLG
ATOGGPLYLVLEAANGNLRLFLRARREPGLDVSFDTCRKPPEQLTKDLVSCAYQV
ARGMEYLAKQCIHRDLDAARNVLVDNVMKLIADFGARUVHNDIYTKKTINGLRPLK
WMAPEALFDRYTHQSDVMSFGVILWEITFLGGSYPGIPVEELFKLKEGHRMDKPA
NCTHDLVMINEGCHAAQSQRPTFKQLVEDLDRLVTSTDEYLDLSAPPEQVSPGQ
DTPSSSSGDDSVFAHDLUPPAPPSSGGST"
join(<5018, .5287,5511, .5576,7131, .7300,7384, .7507,
7599, .7789,9456, .9600,10094, .10284,10588, .10733,
11119, .11240,11323, .11433,11514, .11704,11815, .11937,
12021, .12091,12310, .12447,12593, .12698,12880, .13026)
/gene="FGFR3"
/notes="alternatively spliced"
/codon_start=3
/product="fibroblast growth factor receptor 3"
/protein_id="AAM22078.1"
/db_xref="GI:20452380"
/translation="VPGPEPGQEOQLVFGSDAVELSCPPPGGPGMGPTVWKDGTGL
VPSERVLPQRLQVLNASHESGAYSCQRQLTORVLCHFSVRVTDAPSSGDDEDED
EAEQTDVDTGAPYWRPERMDKLLAVPAANTVRFCPAAGNTRQITLTDVLERSPHRIL
BHGRIKLRHQMSLVMSVSPSDEGRNCTCVENKFGSIRQTYTLDVLERSPHRIL
QAGPANTQAVLSEVFEHCYVSDAQPHIQMLKHEVNGSKVGPDTPVTYVLKTAG
ANTTDKEVLVLSLNVITFEDAGYVSDAQPHIGFHHASMLVLPASEELVEADEAGS
VTAGILSYGVGFLLVIAAVALTCLRLSPPKGLGSPTVHKISRFPLKROVLESNA
SNSNTPPLVARIARLSSGEGPTLVANSELELPADPKWELSRALTGLKPGEGCGQVQ
MAEALGIDKDRAPQVTVAVKMLKDDATDKDLDLSEMEMMKMILGHKNTINLLGAC
TOGGPLYLVLEAANGNLRLFLRARREPGLDVSFDTCRKPPEQLTKDLVSCAYQV
GWELYSKQCIHRDLDAARNVLVDNVMKLIADFGARUVHNDIYTKKTINGRLPVLKWM
APEALFDRYTHQSDVMSFGVILWEITFLGGSYPGIPVEELFKLKEGHRMDKPCAN
THDLVMINEGCHAAQSQRPTFKQLVEDLDRLVTSTDEYLDLSAPPEQVSPGQDGT
PSSSSGDDSVFAHDLUPPAPPSSGGST"
5018, .5287
/gene="FGFR3"
/number=3
5511, .5576
/gene="FGFR3"
/number=4
7131, .7300
/gene="FGFR3"
/number=5
7384, .7507
/gene="FGFR3"
/number=6
7599, .7789
/gene="FGFR3"
/number=7
8678, .8828
/gene="FGFR3"
/notes="alternatively spliced"

```

```

DEFINITION Sequence 1364 from Patent WO03008583.
ACCESSION AX695737
VERSION AX695737.1 GI:29418891
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Morris,D.W. and Engelhard,E.K.
TITLE Novel compositions and methods for cancer
JOURNAL Patent: WO 03008583-A 1364 30-JAN-2003;
Sagres Discovery (US)
FEATURES
    source
        1..33352
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 385; DB 6; Length 33352;
Best Local Similarity 100.0%; Pred. No. 3.3e-87;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAGCTGGTATTTCATACAAATTCCTCAATTCGTGTGTCGCCAGGCA 60
DB 24436 GACTTCAAGCAGCTGGTATTTCATACAAATTCCTCAATTCGTGTGTCGCCAGGCA 24495
QY 61 GGAGACGGTTTCCAGGGAGGGCCGCCCTGTGTGACGGTCCGATGTTATTAGATGTT 120
DB 24496 GGAGACGGTTTCCAGGGAGGGCCGCCCTGTGTGACGGTCCGATGTTATTAGATGTT 24555
QY 121 ACAAGTTTATATATCTATATATATAATTAATTTAGTTTTCACAGATGTTATTTGTTGT 180
DB 24556 ACAAGTTTATATATCTATATATATAATTTAGTTTTCACAGATGTTATTTGTTGT 24615
QY 181 AGACTTAAACACTTCTTACGCAATGCTCTAGAGTTTATAGCTGACGCTACCTTTCA 240
DB 24616 AGACTTAAACACTTCTTACGCAATGCTCTAGAGTTTATAGCTGACGCTACCTTTCA 24675
QY 241 AGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTACTGGCCCTGAG 300
DB 24676 AGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTGTTCTGTACTGTACTGGCCCTGAG 24735
QY 301 TCTGGCAGCTTCCTCTGCTGCTGCGCAGGCCATGCTCAGGCTGGTCTCTCTTGGG 360
DB 24736 TCTGGCAGCTTCCTCTGCTGCTGCGCAGGCCATGCTCAGGCTGGTCTCTCTTGGG 24795
QY 361 GCCCAGTCATGTCGCCAGAGGTG 385
DB 24796 GCCCAGTCATGTCGCCAGAGGTG 24820
RESULT 7
AC016773
LOCUS AC016773 Homo sapiens BAC clone RP11-572017 from 4, complete sequence.
DEFINITION AC016773 189223 bp DNA linear PRI 21-FEB-2002
ACCESSION AC016773
VERSION AC016773.8 GI:18250163
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 189223)
AUTHORS Nguyen,C., Haakenson,W. and Spalding,L.
TITLE The sequence of Homo sapiens BAC clone RP11-572017

```

```

JOURNAL REFERENCE
3 (Bases 1 to 189223)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 4 (Bases 1 to 189223)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2002) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 5 (Bases 1 to 189223)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jan 19, 2002 this sequence version replaced gi:17647088.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0572017
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osceogawa K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-572017; actual end is at base position 189223 of RP11-572017.

The sequence from 143509 to 143570 is covered only by a PCR product of clone DNA.

```

FEATURES
    source
        1..189223
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4"
        /clone="RP11-572017"
        /clone_lib="RPCI-11"
    misc_feature
        1249..2212

```



```

* 15648 15747: gap of 100 bp
* 15748 16499: contig of 752 bp in length
* 16500 16599: gap of 100 bp
* 16600 17345: contig of 746 bp in length
* 17346 17445: gap of 100 bp
* 17446 18126: contig of 681 bp in length
* 18127 18226: gap of 100 bp
* 18227 18979: contig of 753 bp in length
* 18980 19079: gap of 100 bp
* 19080 19806: contig of 727 bp in length
* 19807 19906: gap of 100 bp
* 19907 20649: contig of 743 bp in length
* 20650 20749: gap of 100 bp
* 20750 21476: contig of 727 bp in length
* 21477 21576: gap of 100 bp
* 21577 22298: contig of 722 bp in length
* 22299 23198: gap of 100 bp
* 23199 23110: contig of 712 bp in length
* 23111 23210: gap of 100 bp
* 23211 23939: contig of 729 bp in length
* 23940 24039: gap of 100 bp
* 24040 24779: contig of 740 bp in length
* 24780 24879: gap of 100 bp
* 24880 25622: contig of 743 bp in length
* 25623 25722: gap of 100 bp
* 25723 26458: contig of 736 bp in length
* 26459 26558: gap of 100 bp
* 26559 27255: contig of 697 bp in length
* 27256 27355: gap of 100 bp
* 27356 28101: contig of 746 bp in length
* 28102 28201: gap of 100 bp
* 28202 28940: contig of 739 bp in length
* 28941 29040: gap of 100 bp
* 29041 29796: contig of 756 bp in length
* 29797 29896: gap of 100 bp
* 29897 30631: contig of 735 bp in length
* 30632 30731: gap of 100 bp
* 30732 31435: contig of 704 bp in length
* 31436 31535: gap of 100 bp
* 31536 32243: contig of 708 bp in length
* 32244 32343: gap of 100 bp
* 32344 33071: contig of 728 bp in length
* 33072 33171: gap of 100 bp
* 33172 33897: contig of 726 bp in length
* 33898 33997: gap of 100 bp
* 33998 34725: contig of 728 bp in length
* 34726 34825: gap of 100 bp
* 34826 35569: contig of 744 bp in length
* 35570 35669: gap of 100 bp
* 35670 36411: contig of 742 bp in length
* 36412 36511: gap of 100 bp
* 36512 37231: contig of 720 bp in length
* 37232 37331: gap of 100 bp
* 37332 38076: contig of 745 bp in length
* 38077 38176: gap of 100 bp
* 38177 38897: contig of 721 bp in length
* 38898 38997: gap of 100 bp
* 38998 39715: contig of 718 bp in length
* 39716 39815: gap of 100 bp
* 39816 40539: contig of 724 bp in length
* 40540 40639: gap of 100 bp
* 40640 41341: contig of 702 bp in length
* 41342 41441: gap of 100 bp
* 41442 42168: contig of 727 bp in length
* 42169 42268: gap of 100 bp
* 42269 43012: contig of 744 bp in length
* 43013 43112: gap of 100 bp
* 43113 43827: contig of 715 bp in length
* 43828 43927: gap of 100 bp
* 43928 44651: contig of 724 bp in length
* 44652 44751: gap of 100 bp
* 44752 45464: contig of 713 bp in length
* 45465 45564: gap of 100 bp

```

```

Query Match      49.1%; Score 189; DB 2; Length 67890;
Best Local Similarity 100.0%; Pred.No. 2.6e-37;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTTGTGTGTCGCCAGGCA 60
Db 1039 GACTTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTTGTGTGTCGCCAGGCA 980

QY 61 GGGAGACGGTTTCCAGGAGGGGGCCGGCCCTGTGTGCAGGTTCCGATCTTATAGATGTT 120
Db 979 GGGAGACGGTTTCCAGGAGGGGGCCGGCCCTGTGTGCAGGTTCCGATCTTATAGATGTT 920

QY 121 ACAAGTTTATATATATCTATATATATAATTTTACAGTTTTCACAGATGATTGTTGT 180
Db 919 ACAAGTTTATATATATCTATATATATAATTTTACAGTTTTCACAGATGATTGTTGT 860

QY 181 AGACTTAAC 189
Db 859 AGACTTAAC 851

```

```

RESULT 10
AX695734          33991 bp      DNA      linear      PAT 31-MAR-2003
LOCUS            Sequence 1361 from Patent WO03008583.
DEFINITION       AX695734
ACCESSION        AX695734
VERSION          AX695734.1 GI:29418888
KEYWORDS
SOURCE           Mus musculus (house mouse)
ORGANISM         Mus musculus
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Morris, D.W. and Engelhard, E.K.
  Novel compositions and methods for cancer
  Patent: WO 03008583-A 1361 30-JAN-2003;
  Sagres Discovery (US)
FEATURES
  source
  1..33991
  /organism="Mus musculus"
  /mol_type="unassigned DNA"
  /db_xref="taxon:10090"
ORIGIN

```

```
Query Match      41.4%; Score 159.2; DB 6; Length 33991;
Best Local Similarity 78.8%; Pred. No. 9.3e-30;
Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATTTCTTAATTTCTGTGTGCCAGGAGGAGACGGTTTCCAGGGAG 80
    |||||
Db 24786 TTTTCATACAGATTATCTTAATTTCTGTATTCAGGAG-GAGCCTGTGCCAGGAA 24844
    |||||
QY 81 GGCGCGCCCTGTGTCCAGGTTCCGATGTTATTAGATGTACAAGTTTATATATCTAT 140
    |||||
Db 24845 GGCGTGGCCCTGCAGAGGTTTCAGATGTTAATAGTTAT-CTGTTTCAAGTTTATCT 24900
    |||||
QY 141 ATATATAATTTATTAGTTTTCACAGAGTATTCTGTAGACTTTAAACACTTCTTAACG 200
    |||||
Db 24901 ATCTATATTTATTAGTTTTCACAGTTGT-TTTCGTGTAGCTTAAACACTTCTTAACG 24959
    |||||
QY 201 AATGCTCTAGAGTTTATAGCTGCTACCTTTCAAGCTTTGGAGGAGGCCGTG 260
    |||||
Db 24960 AGTGCTCTAGACTTTTATAGCTGCTACCTTTCAAGCTTTGGAGACAGTGCTG 25019
    |||||
QY 261 AATTC-AGTTGGTCTGCTGTACTCTTACTGGCCCTGAGCTGGGCGAGCTGCTCCCTG 319
    |||||
Db 25020 AATGCAATTTTGTACTTTTGTACTGTCTACTGGCCCTAGGCTTGGTGGCTGCTCCCTG 25079
    |||||
QY 320 CTTGCC 325
    |||||
Db 25080 CTTGTC 25085
    |||||

RESULT 11
AC079504/c
LOCUS
DEFINITION Mus musculus clone RP23-256K13, WORKING DRAFT SEQUENCE, 19
unordered pieces.
ACCESSION AC079504
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 201973)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201973)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1835429
Center clone name: RPCI-23_256K13
-----
Summary Statistics
Consensus quality: 186841 bases at least Q40
Consensus quality: 195099 bases at least Q30
Consensus quality: 196746 bases at least Q20
Estimated insert size: 208000; agarose-fp estimation
Quality coverage: 6.23 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.47 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

```
* as soon as it is available and the accession number will
* be preserved.
1
1006: contig of 1006 bp in length
1106: gap of unknown length
2407: contig of 1301 bp in length
2507: gap of unknown length
2508: contig of 1615 bp in length
4123: gap of unknown length
4223: contig of 1205 bp in length
5427: gap of unknown length
5527: contig of 3093 bp in length
5528: gap of unknown length
8621: contig of 1072 bp in length
8721: gap of unknown length
9792: contig of 2097 bp in length
9793: gap of unknown length
11989: contig of 2233 bp in length
12090: gap of unknown length
14322: contig of 2056 bp in length
14423: gap of unknown length
16479: contig of 3105 bp in length
16579: gap of unknown length
19684: contig of 11549 bp in length
19784: gap of unknown length
31333: contig of 9890 bp in length
31433: gap of unknown length
41323: contig of 12358 bp in length
41423: gap of unknown length
53780: contig of 20886 bp in length
53781: gap of unknown length
53881: contig of 18207 bp in length
74767: gap of unknown length
74866: contig of 28180 bp in length
93074: gap of unknown length
93174: contig of 25483 bp in length
121354: gap of unknown length
121453: contig of 25483 bp in length
146936: gap of unknown length
147036: contig of 28265 bp in length
173301: gap of unknown length
173401: contig of 28572 bp in length.
173402: gap of unknown length
173403: contig of 28572 bp in length.

FEATURES
Location/Qualifiers
1..201973
/organism="Mus musculus"
/db_type="genomic DNA"
/mol_xref="taxon:10090"
/clone="RP23-256K13"
/clone_lib="RPCI mouse BAC library 23"

ORIGIN
Query Match      41.4%; Score 159.2; DB 2; Length 201973;
Best Local Similarity 78.8%; Pred. No. 1.1e-29;
Matches 241; Conservative 0; Mismatches 58; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATTTCTTAATTTCTGTGTGCCAGGAGGAGACGGTTTCCAGGGAG 80
    |||||
Db 149506 TTTTCATACAGATTATCTTAATTTCTGTATTCAGGAG-GAGCCTGTGCCAGGAA 149448
    |||||
QY 81 GGCGCGCCCTGTGTCCAGGTTCCGATGTTATTAGATGTACAAGTTTATATATCTAT 140
    |||||
Db 149447 GGCGTGGCCCTGCAGAGGTTTCAGATGTTAATAGTTAT-CTGTTTCAAGTTTATCT 149392
    |||||
QY 141 ATATATAATTTATTAGTTTTCACAGATGTATTGTTGTAGACTTAAACACTTCTTAACG 200
    |||||
Db 149391 ATCTATATTTATTAGTTTTCACAGTTGT-TTTCGTGTAGGCTTAAACACTTCTTAACG 149333
    |||||
QY 201 AATGCTCTAGAGTTTATAGCTGCTACCTTTCAAGCTTTGGAGGAGGCCGTG 260
    |||||
Db 149332 AGTGCTCTAGACTTTTATAGCTGCTACCTTTCAAGCTTTGGAGACAGTGCTG 149273
    |||||
QY 261 AATTC-AGTTGGTCTGCTGTACTCTTACTGGGCGCTGAGCTGGGCGAGCTGCTCCCTG 319
    |||||
Db 149272 AATGCAATTTTGTACTTTTGTACTGTCTACTGGGCGCTAGGCTTGGTGGCTGCTCCCTG 149213
    |||||
QY 320 CTTGCC 325
    |||||
```


FEATURES	Location/Qualifiers	
	1..1848	/organism="Mus musculus"
source	/mol_type="genomic DNA"	
	/strain="129"	
	/db_xref="taxon:10090"	
	/clone="35F1"	
	/cell_type="embryonic stem cell"	
	/clone_lib="sCos-1"	
gene	join(L42116.1:207..211,L42117.1:1..530,L42118.1:1..5163, L42119.1:1..296,L42120.1:1..1474,L42121.1:1..102, L42122.1:1..108,L42123.1:1..484,L42124.1:1..364, L42125.1:1..341,L42126.1:1..90,L42127.1:1..93, L42128.1:1..129,L42129.1:1..151,L42130.1:1..244, L42131.1:1..165,1..1811)	
	/gene="FGFR3"	
exon	<1..7	
intron	/gene="FGFR3"	
	/number=18	
exon	/gene="FGFR3"	
	/number=18	
polyA signal	/gene="FGFR3"	
	/number=19	
polyA_site	1812..1817	
	1831	
ORIGIN	/note="The cdna and genomic sequence diverge at this pt."	
	Query Match 40.9%; Score 157.6; DB 10; Length 1848; Best Local Similarity 78.4%; Pred. No. 1.8e-29; Matches 240; Conservative 0; Mismatches 59; Indels 7; Gaps 4;	
Qy	21	TTTTTACAAATCTTCTAATTCGTGTGTCCTCCAGGAGGAGACGGTTCCAGGGAG 80
Db	1397	TTTTTCACAGATATCTAATTCGTATGATTCAGGAG-GAGCCTGTGCCAGGGAA 1455
Qy	81	GGCGCGCCCTGTGTGCAGTTCCGATGTTATAGATGTTTCAAGTTTATATATCTAT 140
Db	1456	GGCGTGCCTCGCAAGAAGTTTCAGATGTTAATAGTTAT----CTGTTCAAGTTTATCT 1511
Qy	141	ATATATAATTTATGATTTTACAGATGTTATTTGTGTAGACTTAAACACTTCTTACGC 200
Db	1512	ATCTATAATTTATGATTTTACAGTTGT-TTTGCTGTAGGCTTAAACACTTCTTATGC 1570
Qy	201	AATGCTTCTAGAGTTTATAGCTGGAGCTGCTACCTTTCAAAGCTTGGAGGAGCCGTG 260
Db	1571	AGTGCTTCTAGACTTTTATAGCTAGACTGCTACCTTTCAAAGCTTGGGAGACAGTGGTG 1630
Qy	261	AATTC-AGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 319
Db	1631	AATGCAATTTTGTACTTTTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1690
Qy	320	CTTGCC 325
Db	1691	CCTGTC 1696
RESULT 14		
BC053056		
LOCUS	BC053056 4156 bp mRNA linear ROD 08-OCT-2003	
DEFINITION	Mus musculus fibroblast growth factor receptor 3, mRNA (cdna clone	
	MGC:62341 IMAGE:5708838), complete cds.	
ACCESSION	BC053056	
VERSION	BC053056.1 GI:31419844	
KEYWORDS	MGC.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 4156)	
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clones="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH-BMAP_EXO"	
	/lab_host="DH10B"	
	/note="Vector: pTX-ASC"	
gene	1..4156	
	/gene="Fgfr3"	
	/note="synonym: HBGFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Fgfr3 protein"	
	/protein_id="AAHS3056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="MVVPACVLVFCVAVAGATSBPPGPQVRVRAAEVPGPEPSCQ	
	EQVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	
	GVYSQHLRTRVLVCVTPADSPGDEGDEEDVAEDTGAPTVTRPRMDKKLLAY	
	PAANTVTRCPAGNTPFTSLWNGKGFGEGRHIGIKLRHQQKQSLVMSVYVPSQ	
	QVAFGSDIVELSCHPPGAGPTGPTVWAKDGTGLVASHRLVGLVQRLQVLAWSHEDA	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/cl	

FEATURES	Location/Qualifiers	
	1..4156	/organism="Mus musculus"
source	/mol_type="mRNA"	
	/strain="C57BL/6"	
	/db_xref="taxon:10090"	
	/clone="MGC:62341 IMAGE:5708838"	
	/tissue_type="Brain, mouse 15.5 dpc"	
	/clone_lib="NIH BMAP_EX0"	
	/lab_host="DH10B"	
	/note="Vector: pYX-ASC"	
gene	1..4156	
	/gene="Pgfr3"	
	/note="synonym: HGFFR"	
	/db_xref="LocusID:14184"	
CDS	/db_xref="MGI:95524"	
	219..2621	
	/codon_start=1	
	/product="Pgfr3 protein"	
	/protein_id="AAH53056.1"	
	/db_xref="GI:31419845"	
	/db_xref="LocusID:14184"	
	/translation="VVVPAVCVLVFCVAVAGATSEPPGQEVRRVRAAEVFPPEFSQ	
	EQVAFSGDITVELSCHPPGGAPTGTWAKDGTGLVASHRILVGPQRLQVFNASHEDA	
	GVYSQHLRLTRVLCHFSVRYTDPASSGDDEGDAEDTGPAYWTRPERMDKLLAV	
	PAANTVPRCPAAGNTPSPISLWLNKGFRGEHRIGGKLRHQQNSLWMSVWFSDRG	

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 114 Row: i Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA GI: 6679786.

Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdopaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdopaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Klausner R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udín, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 4156)
 Strausberg, R.
 Direct Submission
 Submitted (02-JUN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdopaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 114 Row: i Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA GI: 6679786.

Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdopaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 CDNA Library Preparation: M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdopaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

NYTCVENKFGSIQTYTLDVLRSPHRPILOAGLPANQTAIILGSDVEFHCKYVSQAQ
PHIQWLKHEVNGSKVDPGTPYVTLKTAGANTDKLEVLISLHNVTFFDAGEYTCCL
ASNPKGSHGSAWLVLPAEELMETDEAGSVYAGVLSYGVVFFLILVAAVILCRL
RSPKGLGSPVTHKVRFPPLKROVLESNSMNSNTPLVRIARLSGEGPVLAVNSE
LELPADPKWELSRTELILGKELGSCFQVVMAEIGIDKORTAKPVTVAVKMLKODA
TKDLSDLVSEMEMMKWIGKXKNIINLLGACTGQGPYLVVEYAAKGNLREFLRARP
PGMDSFACRLEPQETKDLVSCAVQVARGMEYLASQKCIHRDLAARNVLVTEQNV
MKIADFGRLARDVHNLDDYKKTNGRLPVKMAPEALFDRVYTHQSDVMSFGVLLWEIF
TLGSPGPIGVESLFLKLEGRHMDKPSCTHDLVIMRECHWAVPSQRTPKQLVE
DLDRILVTSTDEYLDLSVPEQYSPGQDTPSSSSGSDSVFTHDLLPGRPTSNCGP
RT"

misc_feature
369..548
/note="IGc2; Region: Immunoglobulin C-2 Type"
/db_xref="CDD:smart00408"

misc_feature
596..938
/note="IG; Region: Immunoglobulin"
/db_xref="CDD:smart00409"

misc_feature
978..1271
/note="IG; Region: Immunoglobulin"
/db_xref="CDD:smart00409"

misc_feature
1614..12444
/note="TyrKc; Region: Tyrosine kinase, catalytic domain"
/db_xref="CDD:smart00219"

ORIGIN

Query Match 40.9%; Score 157.6; DB 10; Length 4156;
Best Local Similarity 78.4%; Pred. No. 2e-29;
Matches 240; Conservative 0; Mismatches 59; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTAATCTGTGTGCCAGGAGAGAGCGTTTCCAGGAG 80
Db 3689 TTTTCATACAGATTATCTAATCTGTATGATTCCAGGAG-GAGCTGTGCCAGGAA 3747

QY 81 GGGCGGCGCTGTGCGAGTCCGATGTTATTAGATTACAGTTTATATATATAT 140
Db 3748 GGGCTGGCGCTGCAAGAGGTTTCAGATTGTAATAGTTAT---CTGTACAAGTTTATCT 3803

QY 141 ATATATAATTTATTGAGTTTACAAAGATGATTGTTGTAGACTTAACACTTCTTACGC 200
Db 3804 ATCATAATTTATTGAGTTTACAAAGTTGT-TTGTGTAGCTTAACACTTCTTATGC 3862

QY 201 AATGCTTCTAGAGTTTATAGCTGGAGCTGCTACTCTTCAAGCTTGGAGGAGCGGTG 260
Db 3863 AGTGCTTCTAGACTTTTATAGCTTGTACTGCTTCAAGCTTGGAGAGAGTGGTA 3922

QY 261 AATTC-AGTTGCTGCTTCTGCTACTGTTACTGGGCGCTGAGTGTGGCAGCTGTCCTTG 319
Db 3923 AATGCAATTTTGTACTTTTGTACTGTCACCTGGGCGCTAGCTTGGTGGCTGTCCCTTG 3982

QY 320 CTGGCC 325
Db 3983 CCTGTC 3988

RESULT 15
MUSMFR3
LOCUS BALB/c fibroblast growth factor receptor 3 (mFR3) mRNA, complete cds.
DEFINITION M81342.1 GI:199144
ACCESSION M81342.1
VERSION M81342.1
KEYWORDS fibroblast growth factor receptor 3; transmembrane protein; tyrosine kinase.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4158)
REFERENCE Ornitz, D.M. and Leder, P.
AUTHORS Ligand specificity and heparin dependence of fibroblast growth factor receptors 1 and 3
TITLE J. Biol. Chem. 267 (23), 16305-16311 (1992)
JOURNAL 92355591
MEDLINE

1379594
COMMENT Original source text: Mus musculus (strain BALB/c, sub species domesticus) (library: Balb/C brain cDNA library in Lambda ZAP, Stratagene, La Jolla, CA) brain cDNA to mRNA.
FEATURES
Location/Qualifiers
source
1..4158
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/tissue_type="brain"
/germline
/tissue_lib="Balb/C brain cDNA library in Lambda ZAP, Stratagene, La Jolla, CA"
gene
1..4158
/gene="mFR3"
227..2632
/gene="mFR3"
/codon_start=1
/product="fibroblast growth factor receptor 3"
/protein_id="AAA39535.1"
/db_xref="GI:199145"
/translation="MVVPACVLVFCVAVVAGATSEPPGQEVVRRAAEVGPBPSSQ
EQVAFSGDTVELSCHPPGAPGTVMKGTGLVASHRLVGPQRLOVNLASHEDA
GVYSCQHLTRRVLCHFSVRVTDAPSSGDDDEGDVAEDTGAPYWRPERMDKLLAV
PAAVTRFCPCAGNTPPSISMLKNGKEPRGEHIGGILKSHQOSSLVMSVSPDRG
NYTCVENKFGSIQTYTLDVLRSPHRPILOAGLPANQTAIILGSDVEFHCKYVSQAQ
PHIQWLKHEVNGSKVDPGTPYVTLKTAGANTDKLEVLISLHNVTFFDAGEYTCCL
ASNPKGSHGSAWLVLPAEELMETDEAGSVYAGVLSYGVVFFLILVAAVILCRL
RSPKGLGSPVTHKVRFPPLKROVLESNSMNSNTPLVRIARLSGEGPVLAVNSE
LELPADPKWELSRTELILGKELGSCFQVVMAEIGIDKORTAKPVTVAVKMLKODA
TKDLSDLVSEMEMMKWIGKXKNIINLLGACTGQGPYLVVEYAAKGNLREFLRARP
PGMDSFACRLEPQETKDLVSCAVQVARGMEYLASQKCIHRDLAARNVLVTEQNV
MKIADFGRLARDVHNLDDYKKTNGRLPVKMAPEALFDRVYTHQSDVMSFGVLLWEIF
TLGSPGPIGVESLFLKLEGRHMDKPSCTHDLVIMRECHWAVPSQRTPKQLVE
DLDRILVTSTDEYLDLSVPEQYSPGQDTPSSSSGSDSVFTHDLLPGRPTSNCGP
PRT"
227..286
/gene="mFR3"

ORIGIN

Query Match 40.9%; Score 157.6; DB 10; Length 4158;
Best Local Similarity 78.4%; Pred. No. 2e-29;
Matches 240; Conservative 0; Mismatches 59; Indels 7; Gaps 4;

QY 21 TTTTCATACAAATCTTCTAATCTGTGTGCCAGGAGAGAGCGTTTCCAGGAG 80
Db 3702 TTTTCATACAGATTATCTAATCTGTATGATTCCAGGAG-GAGCTGTGCCAGGAA 3760

QY 81 GGGCGGCGCTGTGTCAGGTTCCGATGTTATTAGATTACAAAGTTTATATATATAT 140
Db 3761 GGGCTGGCGCTGCAAGAAGTTTCAGATTGTAATAGTTAT---CTGTTCAGATTATCT 3816

QY 141 ATATATAATTTATTGAGTTTACAAAGATGATTGTTGTAGACTTAACACTTCTTACGC 200
Db 3817 ATCTATAATTTATTGAGTTTACAAAGTTGT-TTGTGTAGGCTTAACACTTCTTATGC 3875

QY 201 AATGCTTCTAGAGTTTATAGCTGGAGCTGCTACTCTTCAAGCTTGGAGGAGCGGTG 260
Db 3876 AGTGCTTCTAGACTTTTATAGCTAGACTGCTACTCTTCAAGCTTGGAGAGAGTGGTG 3935

QY 261 AATTC-AGTTGCTGCTTCTGCTACTGTTACTGGGCGCTGAGTGTGGGAGAGTGTCCCTTG 319
Db 3936 AATGCAATTTTGTACTTTTGTACTGTCACCTGGGCGCTAGCTTGGGTGGCTGTCCCTTG 3995

QY 320 CTGGCC 325
Db 3996 CCTGTC 4001

Search completed: September 22, 2004, 11:50:38
Job time : 1790 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2004, 10:57:41 ; Search time 2467 Seconds
(without alignments)
4660.292 Million cell updates/sec

Title: M64347

Perfect score: 385

Sequence: 1 GACTTCAGGACGCTGGTA.....GTGCTAGTGGCCAGAGGTG 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estnu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pin:**

20: em_gss_fun:**

21: em_gss_mam:**

22: em_gss_mus:**

23: em_gss_pro:**

24: em_gss_rod:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
C 1	385	100.0	471	10	BF059608	BF059608 7k63d12.x
C 2	385	100.0	476	9	AI343936	AI343936 qp42d08.x
C 3	385	100.0	479	9	AI078769	AI078769 cz34b04.x
C 4	385	100.0	481	10	AW206005	AW206005 UI-H-B11-

RESULT 1

BF059608/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

ALIGNMENTS

C 5	385	100.0	489	9	AI624729	AI624729 ts71b08.x
C 6	385	100.0	492	9	AI080060	AI080060 cz32b04.x
C 7	385	100.0	500	10	BG059722	BG059722 naf52f12.
C 8	385	100.0	537	9	AA410992	AA410992 zv03a11.s
C 9	385	100.0	544	9	AA411000	AA411000 zv03b10.s
C 10	385	100.0	556	12	BQ021570	BQ021570 UI-H-B11-
C 11	385	100.0	575	9	AL043615	AL043615 DKFp434H
C 12	385	100.0	594	14	CA445194	CA445194 UI-H-E11-
C 13	385	100.0	620	9	AA873489	AA873489 oh77h01.s
C 14	385	100.0	649	10	BE466124	BE466124 hy10e11.s
C 15	385	100.0	671	9	AI924133	AI924133 wn64a04.x
C 16	385	100.0	675	9	AI927305	AI927305 wo89d03.x
C 17	385	100.0	680	14	CE851497	CE851497 UI-CF-EN1
C 18	385	100.0	682	9	AA913622	AA913622 ol38d10.s
C 19	385	100.0	682	13	BU683104	BU683104 UI-CF-EC1
C 20	385	100.0	683	12	BM670295	BM670295 UI-E-DW1-
C 21	385	100.0	693	13	BQ446455	BQ446455 UI-H-EV1-
C 22	385	100.0	700	9	AI521743	AI521743 t182d05.x
C 23	385	100.0	713	12	BM982604	BM982604 UI-CF-EN1
C 24	385	100.0	744	12	BM989149	BM989149 UI-CF-EN0
C 25	385	100.0	744	13	BQ574240	BQ574240 UI-H-EZ1-
C 26	385	100.0	770	9	AI057095	AI057095 oz23f02.x
C 27	383.4	99.6	523	10	AW204106	AW204106 UI-H-B11-
C 28	381.8	99.2	589	10	AW268519	AW268519 xv51b10.x
C 29	381.8	99.2	643	9	AI796742	AI796742 wa14d03.x
C 30	381.8	99.2	661	9	AI598047	AI598047 ts05h09.x
C 31	381	99.0	544	9	AI332806	AI332806 qp98d03.x
C 32	377.6	98.1	553	9	AW009550	AW009550 ws84c02.x
C 33	377.6	98.1	721	9	AW074098	AW074098 xb07f11.x
C 34	374	97.1	466	10	BF223012	BF223012 7g27d09.x
C 35	373.8	97.1	511	9	AA909405	AA909405 om31g08.s
C 36	372.8	96.8	1037	13	BX377171	BX377171 BX377171
C 37	371	96.4	601	14	CD674871	CD674871 f816c10.y
C 38	370.4	96.2	691	14	CK005545	CK005545 AGENCOURT
C 39	369.8	96.1	475	9	AA928957	AA928957 o031f10.s
C 40	369.8	96.1	504	9	AI199931	AI199931 qf89e04.x
C 41	367	95.3	483	9	AI655144	AI655144 wb54b09.x
C 42	365.8	95.0	627	10	AW779920	AW779920 hn90g11.x
C 43	362.6	94.2	542	9	AA913131	AA913131 om53d10.s
C 44	360.2	93.6	554	9	AI690405	AI690405 tx87c08.x
C 45	356.6	92.7	424	9	AA419620	AA419620 zv04b08.s

BF059608 7k63d12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3480238 3',
linear EST 16-OCT-2000
mRNA sequence.

BF059608.1 GI:10813504

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 471)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

R. Emmer-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 454.
Location/Qualifiers

1. 471
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3480238"
/tissue type="pooled germ cell tumors"
/lab host="DH10B"
/clone lib="NCI CGAP GC6"
/note=Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 10; Length 471;
Best Local Similarity 100.0%; Pred. No. 7e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTGTCGCCAGGCA 60
DB 458 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTGTCGCCAGGCA 399

QY 61 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTT 120
DB 398 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTT 339

QY 121 ACAAGTTTATATATATCTATATATATAATTTATTTAGAGTTTTCACAGATGTTTGTGT 180
DB 338 ACAAGTTTATATATATCTATATATATAATTTATTTAGAGTTTTCACAGATGTTTGTGT 279

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTACCTTTCA 240
DB 278 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTACCTTTCA 219

QY 241 AAGCTTGGAGGAAGCCCGTGAATTCAGTTGGTTCGTTCTGTTCTGTTCTGTTCTGTTCTG 300
DB 218 AAGCTTGGAGGAAGCCCGTGAATTCAGTTGGTTCGTTCTGTTCTGTTCTGTTCTGTTCTG 159

QY 301 TCTGGGACGCTGCCCTTGTCTGCTGCAGGCGCATGCTCAGGGTGGTCTCTTCTTGGG 360
DB 158 TCTGGGACGCTGCCCTTGTCTGCTGCAGGCGCATGCTCAGGGTGGTCTCTTCTTGGG 99

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 98 GCCCAGTGCATGGTGGCCAGAGGTG 74

RESULT 2
AI343936/c
LOCUS
DEFINITION
qp42d08.x1 NCI CGAP C08 Homo sapiens cDNA clone IMAGE:1925679 3,
similar to gb:M64347 FBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D.; Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennox, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
Location/Qualifiers

1. 476
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925679"
/tissue type="adenocarcinoma"
/lab host="DH10B"
/clone lib="NCI CGAP C08"
/note=Organ: colon; Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p77T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 476;
Best Local Similarity 100.0%; Pred. No. 7e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTGTCGCCAGGCA 60
DB 467 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTGTCGCCAGGCA 408

QY 61 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTT 120
DB 407 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTT 348

QY 121 ACAAGTTTATATATCTATATATATAATTTATTTAGAGTTTTCACAGATGTTTGTGT 180
DB 347 ACAAGTTTATATATCTATATATATAATTTATTTAGAGTTTTCACAGATGTTTGTGT 288

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTACCTTTCA 240
DB 287 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTACCTTTCA 228

QY 241 AAGCTTGGAGGAAGCCCGTGAATTCAGTTGGTTCGTTCTGTTCTGTTCTGTTCTGTTCTG 300
DB 227 AAGCTTGGAGGAAGCCCGTGAATTCAGTTGGTTCGTTCTGTTCTGTTCTGTTCTGTTCTG 168

QY 301 TCTGGGACGCTGCCCTTGTCTGCTGCAGGCGCATGCTCAGGGTGGTCTCTTCTTGGG 360
DB 167 TCTGGGACGCTGCCCTTGTCTGCTGCAGGCGCATGCTCAGGGTGGTCTCTTCTTGGG 108

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 107 GCCCAGTGCATGGTGGCCAGAGGTG 83

RESULT 3
AI078769/c
LOCUS
DEFINITION
oz34504.x1 Soares NHMPu S1 Homo sapiens cDNA clone IMAGE:1677199
3', similar to gb:M64347 FBROBLAST GROWTH FACTOR RECEPTOR 3
PRECURSOR (HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D.; Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennox, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
Location/Qualifiers

1. 476
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1925679"
/tissue type="adenocarcinoma"
/lab host="DH10B"
/clone lib="NCI CGAP C08"
/note=Organ: colon; Vector: p77T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p77T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 476;
Best Local Similarity 100.0%; Pred. No. 7e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTGTCGCCAGGCA 60
DB 467 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCCTGCTGTGTCGCCAGGCA 408

QY 61 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTT 120
DB 407 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTT 348

QY 121 ACAAGTTTATATATCTATATATATAATTTATTTAGAGTTTTCACAGATGTTTGTGT 180
DB 347 ACAAGTTTATATATCTATATATATAATTTATTTAGAGTTTTCACAGATGTTTGTGT 288

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTACCTTTCA 240
DB 287 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTACCTTTCA 228

QY 241 AAGCTTGGAGGAAGCCCGTGAATTCAGTTGGTTCGTTCTGTTCTGTTCTGTTCTGTTCTG 300
DB 227 AAGCTTGGAGGAAGCCCGTGAATTCAGTTGGTTCGTTCTGTTCTGTTCTGTTCTGTTCTG 168

QY 301 TCTGGGACGCTGCCCTTGTCTGCTGCAGGCGCATGCTCAGGGTGGTCTCTTCTTGGG 360
DB 167 TCTGGGACGCTGCCCTTGTCTGCTGCAGGCGCATGCTCAGGGTGGTCTCTTCTTGGG 108

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 107 GCCCAGTGCATGGTGGCCAGAGGTG 83

RESULT 3
AI078769/c
LOCUS
DEFINITION
oz34504.x1 Soares NHMPu S1 Homo sapiens cDNA clone IMAGE:1677199
3', similar to gb:M64347 FBROBLAST GROWTH FACTOR RECEPTOR 3
PRECURSOR (HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m3 fwd. Et from Amersham
High quality sequence stop: 444.
FEATURES
Location/Qualifiers
1..479
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1677199"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHPu S1"
/notes="Organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site: 1: Not 1;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NHPu, and fetal heart NBH19w) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. Clones 260232-265223,
340488-345479, and 484488-489479."
ORIGIN
Query Match 100.0%; Score 385; DB 9; Length 479;
Best Local Similarity 100.0%; Pred. No. 7e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTCGTGTGTGTCCTCCAGGCA 60
DB 474 GACTTCAAAGCAAGCTGGTATTTTCATACAAATTTCTTAATTCGTGTGTGTCCTCCAGGCA 415
QY 61 GCGAGACGGTTCCAGGAGGGCGGCGCTGTGTCAGGTTCCGATGTTATAGATGTT 120
DB 414 GCGAGACGGTTCCAGGAGGGCGGCGCTGTGTCAGGTTCCGATGTTATAGATGTT 355
QY 121 ACAAGTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 180
DB 354 ACAAGTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 295
QY 181 AGACTTTAACTTCTTACGAATGCTTCTAGAGTTTATAGCTGACGTGACCTTTCA 240
DB 294 AGACTTTAACTTCTTACGAATGCTTCTAGAGTTTATAGCTGACGTGACCTTTCA 235
QY 241 AAGCTTGGAGGGAAGCGTCAATTCAGTTGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 300
DB 234 AAGCTTGGAGGGAAGCGTCAATTCAGTTGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 175
QY 301 TCTGGGAGCTGTCCTTGTCTGCTGTCAGGGCCATCGGTCAGGGTGGTCTCTCTTTGGG 360
DB 174 TCTGGGAGCTGTCCTTGTCTGCTGTCAGGGCCATCGGTCAGGGTGGTCTCTCTTTGGG 115
QY 361 GCCAGTGCATGCTGGCCAGAGGTG 385
DB 114 GCCAGTGCATGCTGGCCAGAGGTG 90
RESULT 4
```

```
AW206005 481 bp mRNA linear EST 02-DEC-1999
LOCUS UI-H-B11-afy-f-05-0-UI-s1 NCI CGAP_Sub3 Homo sapiens cDNA clone
DEFINITION IMAGE:2723480 3', mRNA sequence.
ACCESSION AW206005
VERSION AW206005.1 GI:6505479
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image.html The following repetitive
elements were found in this cDNA sequence: 328-356,
>AT_rich#Low complexity
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2723480"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Sub3"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; The
NCI CGAP Sub3 library is a subtracted library derived from
the NCI CGAP Sub1 library, which is a subtracted library
derived from B1. B1 constitutes a mixture of 21
normalized or subtracted NCI CGAP libraries:
NCI CGAP Co4, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
NCI CGAP Br2, NCI CGAP Co8, NCI CGAP Cui1, NCI CGAP Lei2,
NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,
NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
NCI CGAP Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911,
1456008-1456775, 1500552-1502855); NCI CGAP Kids pool 1
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones
1323912-1325831, 1471368-1472903, 1492104-1493255);
NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
Clones 1414920-1417991, 1520904-1522439); NCI CGAP GC4
pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clones 1257096-1258631, 1469064-1470983, LLAM 2457-2459,
1475592-1476743); NCI CGAP Pr22 pool 1
2758-2759, 3062-3068 (IMAGE Clones 985608-986759,
1101192-1101959, 1217928-1220615); NCI CGAP Co10 pool 1
LLAM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255,
1144584-1145351). Subtraction was performed as previously
described [Bonaldo, Lennon & Soares (1996): Normalization
and Subtraction: Two Approaches to Facilitate Gene
Discovery. Genome Research 6, 791-806.].
TAG TISSUE=lung
TAG LIB=NCI CGAP_Lu19
TAG_SEQ=GACGCG
```

ORIGIN

Query Match 100.0%; Score 385; DB 10; Length 481;
 Best Local Similarity 100.0%; Pred. No. 7e-86;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGTTATTTTCATACAAATCTTCTAATTCCTGCTGTGCCAGGCA 60
 DB 481 GACTTCAAAGCAAGCTGTTATTTTCATACAAATCTTCTAATTCCTGCTGTGCCAGGCA 422

QY 61 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTCAGGTTCCGATGTCGATGTTATAGATTT 120
 DB 421 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTCAGGTTCCGATGTCGATGTTATAGATTT 362

QY 121 ACAAGTTTATATATCTAT 180
 DB 361 ACAAGTTTATATATCTAT 302

QY 181 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 240
 DB 301 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 242

QY 241 AAGCTTGGAGGAGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 DB 241 AAGCTTGGAGGAGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182

QY 301 TCTGGGACGCTGTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 DB 181 TCTGGGACGCTGTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
 DB 121 GCCCAGTGCATGGTGGCCAGAGGTG 97

RESULT 5

AI624729/c
 LOCUS ts71b08.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2236695 3'
 DEFINITION similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
 (HUMAN);, mRNA sequence.

ACCESSION AI624729
 VERSION AI624729.1 GI:4649660
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 489)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Ronaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 249 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 448
 POLYA=No.

FEATURES
 source Location/Qualifiers
 1..489
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /dd_xref="taxon:9606"

/clone="IMAGE:2236695"
 /tissue type="pooled germ cell tumors"
 /lab host="DH10B"
 /clone_lib="NCI CGAP GC6"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
 from the normalized library NCI CGAP GC4 was prepared, and
 ss circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones from the same library (clones 1257096-1258631,
 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Ronaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 489;
 Best Local Similarity 100.0%; Pred. No. 7e-86;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGTTATTTTCATACAAATCTTCTAATTCCTGCTGTGCCAGGCA 60
 DB 458 GACTTCAAAGCAAGCTGTTATTTTCATACAAATCTTCTAATTCCTGCTGTGCCAGGCA 399

QY 61 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTCAGGTTCCGATGTCGATGTTATAGATTT 120
 DB 398 GGGAGACGGTTTCCAGGAGGGCGGCCCTGTGTCAGGTTCCGATGTTATAGATTT 339

QY 121 ACAAGTTTATATATCTAT 180
 DB 338 ACAAGTTTATATATCTAT 279

QY 181 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 240
 DB 278 AGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTTGGACTGCTACCTTTCA 219

QY 241 AAGCTTGGAGGAGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 DB 218 AAGCTTGGAGGAGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 159

QY 301 TCTGGGACGCTGTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 DB 158 TCTGGGACGCTGTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 99

QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
 DB 98 GCCCAGTGCATGGTGGCCAGAGGTG 74

RESULT 6

AI080060/c
 LOCUS O232b04.x1 Soares total_fetus Nb2HP8.9w Homo sapiens cDNA clone
 DEFINITION IMAGE:1677007 3' similar to gb:M64347 FIBROBLAST GROWTH FACTOR
 RECEPTOR 3 PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION AI080060
 VERSION AI080060.1 GI:3416311
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 492)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 386 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 332.

Email: cgapbs@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLML, send email to: info@image.llnl.gov

Seq primer: -400P from Gibco
High quality sequence stop: 435.
Location/Qualifiers
1. 500
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4147750"
/tissue_type="glioblastoma without EGFR amplification"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Brn65"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.77 kb. Constructed by Life Technologies."

ORIGIN

	Query Match	100.0%;	Score 385;	DB 10;	Length 500;
	Best Local Similarity	100.0%;	Pred. No. 7e-86;		
	Matches 385;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTTAATTCGTGTGTCGCCAGCA	60		
Db	458	GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTTAATTCGTGTGTCGCCAGCA	399		
QY	61	GGGAGACGGTTCCAGGGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT	120		
Db	398	GGGAGACGGTTCCAGGGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT	339		
QY	121	ACAAAGTTTATATATATCTATATATATAATTAATTAATGAGTTTTTCAAGATGTTATTGTTGT	180		
Db	338	ACAAAGTTTATATATATCTATATATATAATTAATTAATGAGTTTTTCAAGATGTTATTGTTGT	279		
QY	181	AGACTTAACACTTCTTACGCAATGCTTCAGAGTTTTTATAGCTGGAGCTGCTACTTCA	240		
Db	278	AGACTTAACACTTCTTACGCAATGCTTCAGAGTTTTTATAGCTGGAGCTGCTACTTCA	219		
QY	241	AAGCTTGGAGGAAGCCGCGAATTCAGTTGGTTCGTTCTGACTGTTTACTGGGCCCTGAG	300		
Db	218	AAGCTTGGAGGAAGCCGCGAATTCAGTTGGTTCGTTCTGACTGTTTACTGGGCCCTGAG	159		
QY	301	TCTGGGAGCTGTCCTCTGCTGTCAGGCGCATGCTCAGGTCGTCCTCTTCTTGGG	360		
Db	158	TCTGGGAGCTGTCCTCTGCTGTCAGGCGCATGCTCAGGTCGTCCTCTTCTTGGG	99		
QY	361	GCCCAAGTCATGCTGGCCAGAGGTG	385		
Db	98	GCCCAAGTCATGCTGGCCAGAGGTG	74		

RESULT 8
AA410992/c
LOCUS
DEFINITION
AA410992 537 bp mRNA linear EST 18-MAY-1997
zv03all.61 Soares NHM/Pu S1 Homo sapiens cDNA clone IMAGE:752540 3', similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR (HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AA410992.1 GI:2070098
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 537)
Hallier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, S., Jost, S.,

Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Scheellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)

TITLE JOURNAL COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: 41ml3 fwd. ET from Amersham

High quality sequence stop: 337.

FEATURES source

Location/Qualifiers
1..537
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5975372"
/db_xref="taxon:9606"
/clone="IMAGE:752540"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPu S1"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 537;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCACAGGCA 60
DB 456 GACTTCAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCACAGGCA 407
QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTCAGGTTCCGAGTTATTAGATGTT 120
DB 406 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTCAGGTTCCGAGTTATTAGATGTT 347
QY 121 ACAAGTTTAT 180
DB 346 ACAAGTTTAT 287
QY 181 AGACTTACACTTTCTTACGCAATCTTCTAGAGTTTATAGCTTCTACCTTTCA 240
DB 286 AGACTTACACTTTCTTACGCAATCTTCTAGAGTTTATAGCTTCTACCTTTCA 227
QY 241 AAGCTTGGAGGGAAGCGGTGAATTCAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 300
DB 226 AAGCTTGGAGGGAAGCGGTGAATTCAGTTGGTTGGTTGGTTGGTTGGTTGGTTGGTTGG 167
QY 301 TCTGGGACGTGTCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 166 TCTGGGACGTGTCCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 107
QY 361 GCCCAGTGCATGGTGGCCAGAGGTG 385
DB 106 GCCCAGTGCATGGTGGCCAGAGGTG 82

AA411000/c
LOCUS
DEFINITION

AA411000 544 bp mRNA linear EST 18-MAY-1997
zv03b10.s1 Soares NHMPu S1 Homo sapiens cDNA clone IMAGE:752539 3'
similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA411000.1 GI:2070106
EST.
Homo sapiens (human)

REFERENCE

1 (bases 1 to 544)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Scheellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: 41ml3 fwd. ET from Amersham

High quality sequence stop: 457.

FEATURES source

Location/Qualifiers
1..544
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5975371"
/db_xref="taxon:9606"
/clone="IMAGE:752539"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares NHMPu S1"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 544;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCACAGGCA 60
DB 456 GACTTCAAGCAAGCTGTATTTTCATACAAATTTCTTAATGCTGTGTCACAGGCA 407
QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTCAGGTTCCGAGTTATTAGATGTT 120
DB 406 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTCAGGTTCCGAGTTATTAGATGTT 347
QY 121 ACAAGTTTAT 180
DB 346 ACAAGTTTAT 287
QY 181 AGACTTACACTTTCTTACGCAATCTTCTAGAGTTTATAGCTTCTACCTTTCA 240
DB 286 AGACTTACACTTTCTTACGCAATCTTCTAGAGTTTATAGCTTCTACCTTTCA 227

```

QY 241 AAGCTTGGAGGAGCGTGAATTGAGTTGTTGCTGTTCTGTTACTGTTACTGTTGGCCCTGAG 300
Db 226 AAGCTTGGAGGAGCGTGAATTGAGTTGTTGCTGTTCTGTTACTGTTACTGTTGGCCCTGAG 167
QY 301 TCTGGGAGTGTCCCTTGTGCTGTCAGGGCCATGGCTCAGGGTGGTCTCTCTTTGGG 360
Db 166 TCTGGGAGTGTCCCTTGTGCTGTCAGGGCCATGGCTCAGGGTGGTCTCTCTTTGGG 107
QY 361 GCCAGTGCATGTGGCCAGAGTG 385
Db 106 GCCAGTGCATGTGGCCAGAGTG 82

RESULT 10
BQ021570/c
LOCUS
DEFINITION
  UI-H-DH1-axg-j-15-0-UI.sl NCI_CGAP_DH1 Homo sapiens cDNA clone
  IMAGE:5828462 3', mRNA sequence.
ACCESSION
  BQ021570
VERSION
  BQ021570.1 GI:19756848
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 556)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps@mail.nih.gov
  Tissue Procurement: Dr. Jose Mercuende
  cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  The following repetitive elements were found in this cDNA
  sequence: 323-351, >AT rich#low_complexity (matched complement)
  Seq primer: M13 FORWARD
  POLYA=Yes.
FEATURES
  source
  1..556
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:5828462"
  /tissue_type="Metastatic Chondrosarcoma"
  /dev_stage="Adult"
  /lab_host="DH10B (Life Technologies)"
  /clone_lib="NCI_CGAP_DH1"
  /notes="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
  modified polylinker; Site 1: EcoR I; Site 2: Not I;
  NCI CGAP DH1 is a normalized cDNA library containing the
  following tissue(s): VS-8 Cell line from Metastatic
  Chondrosarcoma in Lung. The library was constructed
  according to Bonaldo, Lennon and Soares, Genome Research,
  6:791-806, 1996. First strand cDNA synthesis was primed
  with an oligo-dT primer containing a Not I site. Double
  stranded cDNA was ligated to an EcoR I adaptor, digested
  with Not I, and cloned directionally into p773-Pac
  vector. The oligonucleotide used to prime the synthesis of
  first-strand cDNA contains a library tag sequence that is
  located between the Not I site and the (dT)18 tail. The
  sequence tag for this library is AGATCATTCG.
  TAG_TISSUE=lung
  TAG_LIB=UI-H-DH1
  TAG_SEQ=AGATCATTCG"
ORIGIN
  Query Match 100.0%; Score 385; DB 12; Length 556;
  Best Local Similarity 100.0%; Pred. No. 7.2e-86;

```

```

Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTTCRAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCCTCCAGGCA 60
Db 476 GACTTCRAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCCTCCAGGCA 417
QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCTGTGTGTCAGGTTCCGATGTTATTAGATTT 120
Db 416 GGGAGACGGTTTCCAGGAGGGGGCCCTGTGTGTCAGGTTCCGATGTTATTAGATTT 357
QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATATAT 180
Db 356 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATATAT 297
QY 181 AGACTTTAACACTTCTTACGCAATGCTTCTAGAGTTTTTATAGCTTGGATGCTTACCTTTCA 240
Db 296 AGACTTTAACACTTCTTACGCAATGCTTCTAGAGTTTTTATAGCTTGGATGCTTACCTTTCA 237
QY 241 AAGCTTGGAGGAGCGTGAATTCAGTTGGTTCGTTCTGTTACTGTTACTGTTGGCCCTGAG 300
Db 236 AAGCTTGGAGGAGCGTGAATTCAGTTGGTTCGTTCTGTTACTGTTACTGTTGGCCCTGAG 177
QY 301 TCTGGGAGCTGTCCCTTGTGCTTGCCTGCAGGGCCATGGCTCAGGTTGGTCTCTTCTTGGG 360
Db 176 TCTGGGAGCTGTCCCTTGTGCTTGCCTGCAGGGCCATGGCTCAGGTTGGTCTCTTCTTGGG 117
QY 361 GCCAGTGCATGTGGCCAGAGTG 385
Db 116 GCCAGTGCATGTGGCCAGAGTG 92

RESULT 11
AL043615/c
LOCUS
DEFINITION
  DKFZp434H1627.s1 434 (synonym: htes3) Homo sapiens cDNA clone
  DKFZp434H1627.3', mRNA sequence.
ACCESSION
  AL043615
VERSION
  AL043615.1 GI:5423002
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 575)
  Bloembergen, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and
  Wiemann, S.
  EST (Bloembergen, et al.)
  Unpublished (1999)
  Contact: MIPS
  MIPS
  Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
  This is the 3' sequence of the clone insert
  Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
  Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
  Sequenced by GBF (National Research Centre for Biotechnology Ltd.,
  Braunschweig/Germany) within the cDNA sequencing consortium of the
  German Genome Project.
  x1 sequence also available.
  This clone (DKFZp434H1627) is available at the RZPD in Berlin.
  Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
  Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
  source
  1..575
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="DKFZp434H1627"
  /tissue_type="testis"
  /dev_stage="adult"
  /lab_host="DH10B"
  /clone_lib="434 (synonym: htes3)"
  /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN

```



```

Query Match      100.0%; Score 385; DB 9; Length 575;
Best Local Similarity 100.0%; Pred. No. 7.2e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGGTATTTTTCATACAAATCTTCTAATTCGTGTGTCGCCAGGCA 60
DB 439 GACTTCAAAGCAAGCTGGTATTTTTCATACAAATCTTCTAATTCGTGTGTCGCCAGGCA 380
QY 61 GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT 120
DB 379 GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT 320
QY 121 ACAAGTTTATATATATATATATATATATATATATATATATATATATATATATATAT 180
DB 319 ACAAGTTTATATATATATATATATATATATATATATATATATATATATATATATAT 260
QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGCTGCTACCTTTCA 240
DB 259 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGCTGCTACCTTTCA 200
QY 241 AGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTCGTTCTGCTACTGTTACTGGGCCCTGAG 300
DB 199 AGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTCGTTCTGCTACTGTTACTGGGCCCTGAG 140
QY 301 TCTGGGCAGCTGTCCTTCTGCTTCTGCTGAGGCGCATGCTCAGGTTGGTCTCTTCTTGGG 360
DB 139 TCTGGGCAGCTGTCCTTCTGCTTCTGCTGAGGCGCATGCTCAGGTTGGTCTCTTCTTGGG 80
QY 361 GCCCAGTGCATGTTGGCCAGAGGTG 385
DB 79 GCCCAGTGCATGTTGGCCAGAGGTG 55

RESULT 12
CA445194 594 bp mRNA linear EST 08-NOV-2002
LOCUS UI-H-E11-azc-f-24-0-UI.s1 NCI CGAP E11 Homo sapiens cDNA clone
DEFINITION UI-H-E11-azc-f-24-0-UI 3', mRNA sequence.
ACCESSION CA445194
VERSION CA445194.1 GI:24809614
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 594)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bent-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 330-358, >AT richLow_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=yes.
FEATURES
source Location/Qualifiers
1..594
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-E11-azc-f-24-0-UI"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_E11"

```

/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI_CGAP_E11 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA, synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTTGGCAC.

TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-E11
TAG_SEQ=AACTTGGCAC"

ORIGIN

```

Query Match      100.0%; Score 385; DB 14; Length 594;
Best Local Similarity 100.0%; Pred. No. 7.3e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGGTATTTTTCATACAAATCTTCTAATTCGTGTGTCGCCAGGCA 60
DB 483 GACTTCAAAGCAAGCTGGTATTTTTCATACAAATCTTCTAATTCGTGTGTCGCCAGGCA 424
QY 61 GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT 120
DB 423 GGGAGACGGTTTCCAGGAGGGGCGGCCCTGTGTCAGGTTCCGATGTTATTAGATGTT 364
QY 121 ACAAGTTTATATATATATATATATATATATATATATATATATATATATATATATAT 180
DB 363 ACAAGTTTATATATATATATATATATATATATATATATATATATATATATATATAT 304
QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGCTGCTACCTTTCA 240
DB 303 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGCTGCTACCTTTCA 244
QY 241 AAGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTCGTTCTGCTACTGTTACTGGGCCCTGAG 300
DB 243 AAGCTTGGAGGGAAGCCGTGAATTCAGTTGGTTCGTTCTGCTACTGTTACTGGGCCCTGAG 184
QY 301 TCTGGGCAGCTGTCCTTCTGCTTCTGCTGAGGCGCATGCTCAGGTTGGTCTCTTCTTGGG 360
DB 183 TCTGGGCAGCTGTCCTTCTGCTTCTGCTGAGGCGCATGCTCAGGTTGGTCTCTTCTTGGG 124
QY 361 GCCCAGTGCATGTTGGCCAGAGGTG 385
DB 123 GCCCAGTGCATGTTGGCCAGAGGTG 99

```

RESULT 13

AA873489/c

LOCUS

DEFINITION

AA873489

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA873489 620 bp mRNA linear EST 29-APR-1998
ch77h01.s1 NCI CGAP Kid3 Homo sapiens cDNA clone IMAGE:1473073 3',
similar to G0:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN);, mRNA sequence.

AA873489

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 620)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert length: 858 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amersham
 High quality sequence stop: 443.
 Location/Qualifiers

FEATURES

1. .620

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1473073"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Kid3"
 /notes="Organ: kidney; Vector: pTT3D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer,
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pTT3 vector. mRNA
 source: 2 pooled kidneys. Library went through one round
 of normalization. Library constructed by Bento Soares and
 M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 7.3e-86;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATTTCTTAATTCGTGTGCCAGGCA 60
 DB 466 GACTTCAAGCAAGCTGGTATTTCATACAAATTTCTTAATTCGTGTGCCAGGCA 407
 QY 61 GGGAGACGGTTCCAGGAGGGGGCCCGCTGTGCGAGGTCGGATGTTATTAGATGT 120
 DB 406 GGGAGACGGTTCCAGGAGGGGGCCCGCTGTGCGAGGTCGGATGTTATTAGATGT 347
 QY 121 ACAAGTTTATATATATCTATATATATAATTTATTAGTTTTCACAGATGTTATTGTGT 180
 DB 346 ACAAGTTTATATATCTATATATATAATTTATTAGTTTTCACAGATGTTATTGTGT 287
 QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACGCTACCTTTCA 240
 DB 286 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACGCTACCTTTCA 227
 QY 241 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTCTGTTACTGTGAGCCCTGAG 300
 DB 226 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTCTGTTACTGTGAGCCCTGAG 167
 QY 301 TCTGGGAGCTGTCCCTTGTCTGCTCAGGGCCATGGCTCAGGGTGTCTCTCTTTGGG 360
 DB 166 TCTGGGAGCTGTCCCTTGTCTGCTCAGGGCCATGGCTCAGGGTGTCTCTCTTTGGG 107
 QY 361 GCCCAGTGCATGTGGGCCAGAGGTG 385
 DB 106 GCCCAGTGCATGTGGGCCAGAGGTG 82

RESULT 14
 BE466124/c
 LOCUS
 DEFINITION
 Hy0611.x1 NCI CGAP G66 Homo sapiens cDNA clone IMAGE:196940 3'
 similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
 (HUMAN); mRNA sequence.
 BE466124
 BE466124.1 GI:9511899
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 649)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov

Seq primer: -40UP from Gibco
 High quality sequence stop: 454.
 Location/Qualifiers

FEATURES

1. .649

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3196940"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP G66"
 /note="Vector: pTT3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA
 from the normalized library NCI CGAP G64 was prepared, and
 ss circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones
 1257096-1258631, 1469064-1470983, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 10; Length 649;
 Best Local Similarity 100.0%; Pred. No. 7.4e-86;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTTCAAGCAAGCTGGTATTTCATACAAATTTCTTAATTCGTGTGCCAGGCA 60
 DB 458 GACTTCAAGCAAGCTGGTATTTCATACAAATTTCTTAATTCGTGTGCCAGGCA 399
 QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGCGAGGTCGGATGTTATTAGATGT 120
 DB 398 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGCGAGGTCGGATGTTATTAGATGT 339
 QY 121 ACAAGTTTATATATCTATATATAATTTATTAGTTTTCACAGATGTTATTGTGT 180
 DB 338 ACAAGTTTATATATCTATATATAATTTATTAGTTTTCACAGATGTTATTGTGT 279
 QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACGCTACCTTTCA 240
 DB 278 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGACGCTACCTTTCA 219
 QY 241 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTCTGTTACTGTGAGCCCTGAG 300
 DB 218 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTCTGTTACTGTGAGCCCTGAG 159
 QY 301 TCTGGGAGCTGTCCCTTGTCTGCTCAGGGCCATGGCTCAGGGTGTCTCTCTTTGGG 360
 DB 158 TCTGGGAGCTGTCCCTTGTCTGCTCAGGGCCATGGCTCAGGGTGTCTCTCTTTGGG 99
 QY 361 GCCCAGTGCATGTGGGCCAGAGGTG 385
 DB 98 GCCCAGTGCATGTGGGCCAGAGGTG 74

RESULT 15

AI924133/c
LOCUS
DEFINITION
wn64a04.x1 NCI CGAP Lu19 Homo sapiens cDNA clone IMAGE:2450190 3',
similar to gb:M64347 FIBROBLAST GROWTH FACTOR RECEPTOR 3 PRECURSOR
(HUMAN); mRNA sequence.

ACCESSION

AI924133

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AI924133.1 GI:5660097
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 671)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 593 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 465.

FEATURES

source

1..671
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2450190"
/tissue type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI_CGAP_Lu19"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 385; DB 9; Length 671;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACITTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCCTCCAGGCA 60
Db 459 GACITTCAAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTCGTGTGTCCTCCAGGCA 400
QY 61 GGGAGAGGTTTCAGGAGGGGGCCCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 120
Db 399 GGGAGAGGTTTCAGGAGGGGGCCCGCTGTGTCAGGTTCCGATGTTATTAGATGTT 340
QY 121 ACAAGTTTATATATATCTATATATATATATTTATTAGTTTTTACAAGATGTTATTGTTG 180
Db 339 ACAAGTTTATATATATCTATATATATATTTATTAGTTTTTACAAGATGTTATTGTTG 280
QY 181 AGACTTAACCTTCTTACGGCAATGCTTCTAGAGTTTTATAGCTGGAGCTGCTACCTTTCA 240
Db 279 AGACTTAACCTTCTTACGGCAATGCTTCTAGAGTTTTATAGCTGGAGCTGCTACCTTTCA 220

QY 241 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTTCTGTACTGTACTGGGCGCTGAG 300
Db 219 AAGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTTCTGTACTGTACTGGGCGCTGAG 160
QY 301 TCTGGGCAGCTGTCCCTTGCCTTGCCTGCAGGGCCATGGCTCAGGTTGGTCTCTTCTGGG 360
Db 159 TCTGGGCAGCTGTCCCTTGCCTTGCCTGCAGGGCCATGGCTCAGGTTGGTCTCTTCTGGG 100
QY 361 GCCCAGTGCATGGTGCCACAGAGTG 385
Db 99 GCCCAGTGCATGGTGCCACAGAGTG 75

Search completed: September 22, 2004, 12:31:55
Job time : 2471 secs

Result No.	Query Match	Score	Query			Description
			Match	Length	ID	
1	41.8	10.9	21234	4	US-09-1010-671-3	Sequence 3, Appli
2	41.8	10.9	21234	4	US-10-1010-884-3	Sequence 3, Appli
3	37.4	9.7	3923	3	US-08-8660-635A-20	Sequence 20, Appli
4	37.4	9.7	3923	4	US-09-2814-476-20	Sequence 20, Appli
C 5	36.6	9.5	4839	4	US-09-3533-332-1	Sequence 1, Appli
C 6	35.8	9.3	6038	3	US-09-3056-639-4	Sequence 4, Appli
C 7	35.8	9.3	6038	4	US-09-5281-160B-2	Sequence 2, Appli
C 8	35.8	9.3	7822	3	US-09-3086-639-1	Sequence 1, Appli
C 9	35.8	9.3	7622	4	US-09-5235-160B-1	Sequence 1, Appli
C 10	35.8	9.3	8791	5	PCF-US96-01735-5	Sequence 5, Appli
C 11	35.8	9.3	9610	4	US-09-5669-921-45	Sequence 5, Appli
12	35.6	9.2	3627	1	US-08-1040-072B-6	Sequence 45, Appli
13	35.6	9.2	3627	1	US-08-3514-413-7	Sequence 6, Appli
14	35.6	9.2	3627	2	US-08-0235-583-7	Sequence 7, Appli
15	35.2	9.1	6368	4	US-10-2040-708-67	Sequence 67, Appli
16	35.2	9.1	640681	4	US-09-7909-988-1	Sequence 1, Appli
C 17	35	8.9	84495	4	US-09-7979-906-3	Sequence 3, Appli
C 18	34.4	8.9	1522	4	US-09-6202-312D-96	Sequence 96, Appl
C 19	34.4	8.9	19333	4	US-10-2040-708-45	Sequence 45, Appli
C 20	34.2	8.5	2614	4	US-09-0040-056-1	Sequence 1, Appli
C 21	34	8.8	1845	4	US-09-1070-532A-2717	Sequence 2717, Ap
22	33.8	8.8	8607	4	US-10-2040-708-71	Sequence 71, Appl
23	33.6	8.7	53332	4	US-09-8018-861-3	Sequence 3, Appli
24	33.4	8.7	8537	4	US-10-2040-708-41	Sequence 41, Appl
25	33.2	8.6	870	4	US-08-9566-171E-879	Sequence 879, App
26	33.2	8.6	3416	2	US-08-4518-822A-15	Sequence 15, Appli
27	33.2	8.6	3416	4	US-08-3323-430-15	Sequence 15, Appli


```

RESULT 5
US-09-353-332-1/c
; Sequence 1, Application US/09353332
; Patent No. 6316697
; GENERAL INFORMATION:
; APPLICANT: Dixon, Richard A.
; APPLICANT: Xia, Yiji
; APPLICANT: Lamb, Christopher
; TITLE OF INVENTION: CONSTITUTIVE DISEASE RESISTANCE (CDRL1)
; TITLE OF INVENTION: GENE AND METHODS OF USE THEREOF
; FILE REFERENCE: SALK2820-1
; CURRENT APPLICATION NUMBER: US/09/353.332
; CURRENT FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092.696
; EARLIER FILING DATE: 1998-07-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4839
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1111)...(2421)
US-09-353-332-1

Query Match          9.5%; Score 36.6; DB 4; Length 4839;
Best Local Similarity 58.9%; Pred. No. 0.22;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 114 AGATGTTACAGTTTATATATATCTATATATATATATATATATGAGTTTTCACAGATGAT 173
Db 3078 ATATGTACAATGTAATTTTATGAATAATAAAATATATTTATTTCTTATAAGATATAT 3019

QY 174 TTGTTGTAGACTTAAACACTTCTTACGCAATGCTTCTAGAGTTTATTA 220
Db 3018 TTATTTTATAGAAACTTAAAGAGAAACCTAATTGAATTATTA 2972

RESULT 6
US-09-305-639-4/c
; Sequence 4, Application US/09305639
; Patent No. 6200778
; GENERAL INFORMATION:
; APPLICANT: Treco, Douglas A.
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Seiden, Richard F.
; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
; FILE REFERENCE: 07236/016001
; CURRENT APPLICATION NUMBER: US/09/305.639
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/084.663
; EARLIER FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 6038
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-305-639-4

Query Match          9.3%; Score 35.8; DB 3; Length 6038;
Best Local Similarity 61.1%; Pred. No. 0.44;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 101 TTCCGATGTTATTAGATGTTACAAAGTTTATATATCTATATATATAATTTTATGAGTTT 160
Db 503 TTCTTTAATGTTTGGTTTATATATATATATATATATATATATATATATATATATATAT 444

QY 161 TTAACAAGATGTTATTTGTTGTAGACTTAACTTCT 195
Db 443 ATATATTTATATAAATATATATATATATATATATATATATATATATATATATATAT 409

```



```

Query Match          9.2%; Score 35.6; DB 1; Length 3627;
Best Local Similarity 47.0%; Pred. No. 0.39;
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
OV      1 GACTTCAAGCAAGCTGGTATTTCATACAAATCTTCTAAATGCTGTGTGCCAGCA 60

```

FEATURE:
NAME/KEY: - 2733..2739
LOCATION: /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: - 2765
LOCATION: /note= "transcription initiation
OTHER INFORMATION: determined by primer extension"
FEATURE:
NAME/KEY: - 2846..2848
LOCATION: /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"
US-08-351-413-7

Query Match 9.2%; Score 35.6; DB 1; Length 3627;
Best Local Similarity 47.0%; Pred. No. 0.39;
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGTATTTTCATACAAATCTTCTAATTCCTGTGTGCCAGGCA 60
DB 1188 GAGAGCAATGCAATTTGCAACAAACACAGGTCCTTCCAAACGGTGGTTTCATTCCA 1247

QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGCGAGTTCGGATGTTATTAGATGTT 120
DB 1248 CAGAACAGGATAGCAACCCAGAGCAACACCGTTCAACATATATATATATATATAT 1307

QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATAT 180
DB 1308 AT 1367

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTGCTAC 234
DB 1368 ATACGGAACCTTAAACACATGAAATCTAAACATTTTCAACCAATCAGAACTAC 1421

RESULT 14
US-09-025-583-7
Sequence 7, Application US/09025583
Patent No. 5977433
GENERAL INFORMATION:
APPLICANT: Williams, Mark
TITLE OF INVENTION: Maintenance of male-sterile plants
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 2046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/025,583
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/351,413
FILING DATE: US 07/899,072
FILING DATE: 12-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/970,849
FILING DATE: 03-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Svensson, Leonard R.

REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 2121-102PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3627 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Oryza sativa
STRAIN: Akihikari
FEATURE:
NAME/KEY: - 1..2845
LOCATION: /label= PT72
OTHER INFORMATION: /note= "sequence comprising anther specific
OTHER INFORMATION: promoter PT72"
FEATURE:
NAME/KEY: - 2733..2739
LOCATION: /label= TATA
OTHER INFORMATION: /note= "TATA Box"
FEATURE:
NAME/KEY: - 2765
LOCATION: /note= "transcription initiation
OTHER INFORMATION: determined by primer extension"
FEATURE:
NAME/KEY: - 2846..2848
LOCATION: /label= ATG
OTHER INFORMATION: /note= "ATG start of translation of rice T72 gene"
US-09-025-583-7

Query Match 9.2%; Score 35.6; DB 2; Length 3627;
Best Local Similarity 47.0%; Pred. No. 0.39;
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 GACTTCAAAGCAAGCTGTATTTTCATACAAATCTTCTAATTCCTGTGTGCCAGGCA 60
DB 1188 GAGAGCAATGCAATTTGCAACAAACACAGGTCCTTCCAAACGGTGGTTTCATTCCA 1247

QY 61 GGGAGACGGTTTCCAGGAGGGGGCCCGCTGTGTGCGAGTTCGGATGTTATTAGATGTT 120
DB 1248 CAGAACAGGATAGCAACCCAGAGCAACACCGTTCAACATATATATATATATATAT 1307

QY 121 ACAAGTTTATATATATCTATATATATATATATATATATATATATATATATATATAT 180
DB 1308 AT 1367

QY 181 AGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGAGTCTGCTAC 234
DB 1368 ATACGGAACCTTAAACACATGAAATCTAAACATTTTCAACCAATCAGAACTAC 1421

RESULT 15
US-10-204-708-67
Sequence 67, Application US/10204708
Patent No. 667731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012

; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 67
; LENGTH: 6368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-67

Query Match 9.1%; Score 35.2; DB 4; Length 6368;
Best Local Similarity 49.0%; Pred. No. 0.7;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 114 AGATGTTACAGTTTATATATCTATATATATATATATATTTTACAGATGTAT 173
Db |||||
4114 AGTTTATTAAGTTTTTATATATGTTGATAGATATATATTTTGTAGGGGGTT 4173
QY 174 TTGTTGTAGACTTAACACTTCTTACGCAATGCTTAGAGTTTTATAGCTGGACTGCTA 233
Db |||||
4174 AGTTTATTTAAATATATAGATTATGTAGTGTGTGMAATTTTAAAGTTATATAGTA 4233
QY 234 CCTTCAAAGCTTGGAGGGAGCCGTAATTCAGTTGGTTCTGTCTGTACTGTACTGGG 293
Db |||||
4234 GGTTTAGGGTTTAAAGTAGAGGAGTCAGTTTATTTCCGAAAGGTTTAGAGTAGTTTTAGG 4293
QY 294 CCTGAGTCTGG 305
Db 4294 TAGTAATTACGG 4305

Search completed: September 22, 2004, 12:33:21
Job time : 78 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:37:05 ; Search time 87.5 Seconds
(without alignments)
2486.418 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAAGCAAGCTGCTA.....GTGCATGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool.p/CANELLA305B/runat.22092004.112334.14487/app.query.fasta_1.583
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CANELLA305B @CGN 1.1.224 @runat.22092004.112334.14487 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94	13.4	561	AAU02978	Angiotensin
2	83	12.0	494	ABG12606	Novel hum
3	83	12.0	504	AA895722	Human pro
4	76	11.0	381	AAW72063	HSV-2 str
5	75	10.9	966	AAU80189	Human TSP
6	72.5	10.3	147	ABB90106	Human pol
7	71	10.3	961	AAU80188	Human TSP
8	71	10.1	1045	ABO07129	Novel hum
9	71	10.1	2273	ADCS1662	Human MEG
10	71	10.1	2778	ADCS1660	Human MEG

11	70.5	10.0	1043	7	ADD04782	Agg04782 Pig RAG-1
12	70.5	10.2	2130	3	AGS50617	Arabidops
13	70.5	10.2	2158	3	AGS50616	Arabidops
14	70.5	10.2	2204	3	AGS50615	Arabidops
15	70.5	10.2	2254	2	AA776949	ACCASE. 2
16	70.5	10.2	2254	2	AA722129	Arabidops
17	70.5	10.2	2254	2	AA40598	Arabidops
18	70	10.1	865	4	ABB64483	Drosophil
19	70	10.1	870	3	AB21252	Rat metal
20	69	10.0	420	6	ABR42500	Coumeryc
21	68.5	9.9	93	5	ABP64921	Human pro
22	68.5	9.8	363	2	AA37863	Amino aci
23	68.5	9.8	391	7	ADC87517	Human GPC
24	68.5	9.8	1040	2	AAW73584	RAG-1 pro
25	68.5	9.9	1186	4	ABB63516	Drosophil
26	68	9.9	168	6	ADA83946	Human POM
27	68	9.9	420	6	ABR42541	Clorobloc
28	67.5	9.6	141	3	ABBS3247	Human col
29	67.5	9.8	396	7	ADB37611	Neural th
30	67.5	9.8	431	2	AA450027	Human dim
31	67.5	9.8	475	2	AA434000	Human dim
32	67.5	9.6	526	5	ABG30696	Murine MD
33	67.5	9.6	527	5	ABB80743	Rat YAK1
34	67.5	9.8	599	6	AAE37108	Human sel
35	67.5	9.8	599	6	ABR42589	Selective
36	67	9.7	326	5	ABBS7350	Mouse isc
37	67	9.5	528	2	AAW64559	Human pro
38	67	9.5	528	4	AAE02011	Human YAK
39	67	9.5	528	7	ABR57642	Human DYR
40	67	9.5	549	3	AAV68781	Amino aci
41	67	9.7	628	7	ADB85485	Human agg
42	67	9.7	752	7	ADB85483	Human agg
43	67	9.7	753	7	ADB85481	Human agg
44	67	9.7	763	7	ADB85492	Human agg
45	67	9.7	930	2	AAW75426	Human agg

ALIGNMENTS

RESULT 1

AAU02978

ID AAU02978 standard; protein; 561 AA.

XX AAU02978;

AC AAU02978;

DT 12-SEP-2001 (first entry)

XX

DE Angiotensin converting enzyme (ACEV) splice variant protein #78.

XX

KW Angiotensin converting enzyme splice variant; ACEV; interleukin 6;

KW granulocyte colony stimulating factor receptor; glucagon; hypertrophy;

KW platelet-derived endothelial cell growth factor; cardiovascular disease;

KW cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;

KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;

KW myocardial infarction; coronary arterial thrombosis; renal disease;

KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;

KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;

KW noncardiotoxic pulmonary granulomatous disease; endothelial abnormality;

XX Homo sapiens.

OS Homo sapiens.

XX WO200136632-A2.

XX

PD 25-MAY-2001.

XX

PF 17-NOV-2000; 2000WO-IL000766.

XX

PR 17-NOV-1999; 99IL-00132978.

PR 10-DEC-1999; 99IL-00133455.

XX (COMP-) COMPUGEN LTD.

XX

PN EPI074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX Claim 8; SEQ ID NO 18594; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX SQ Sequence 504 AA;
Alignment Scores:
Pred. No.: 0.254 Length: 504
Score: 83.00 Matches: 26
Percent Similarity: 47.14% Conservative: 7
Best Local Similarity: 37.14% Mismatches: 19
Query Match: 12.03% Indels: 18
DB: 4 Gaps: 3
M64347 (1-385) x AAB95722 (1-504)
QY 364 GGGCCCCAGAGAGACCCACCTGACCGCCCTGACGAGCAGGACAGCTGCC 305
DB 425 GlyProGlnGluArgThrProCys-----GlyProAlaLeuArgGluGlyGluAla 442
QY 304 CAGACTCAGGGCCCACTAAGTACAGAAC-----GAACCA 269
DB 443 AlaValGluGlyProCysProSerGlnGluSerLeuSerGlnGluAsnProGluPro 462
QY 268 ACTGAATTCACGGCTTCC-----CTCAGCTTTGAAAGGT 233
DB 463 ThrGluAspGluArgSerGluGluLysGlyGlyValGluValLeuGluSerCysGlnGly 482

QY 232 AGCAGTCCAGGCTATATAAACTCTAGAAGCA 203
DB 483 SerSerAsnGlyAlaGlnAspGlnGluAla 492
RESULT 4
AAW72063
ID AAW72063 standard; protein; 381 AA.
XX AAW72063;
XX 18-DEC-1998 (first entry)
XX HSV-2 strain SB5 Contig ID 93 ORF#1 protein.
XX HSV-2 strain SB5; immunological response induction; therapy;
XX antiviral identification; viral protein inhibitor.
XX Herpes simplex virus 2.
XX WO9820016-A1.
XX 14-MAY-1998.
XX 31-OCT-1997; 97WO-US020016.
XX 04-NOV-1996; 96US-0030279P.
XX 09-JUN-1997; 97US-0049018P.
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX Esser KM, Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;
XX Leary JJ;
XX WPI; 1998-286847/25.
XX N-FSDB; AAV62147.
XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in mammal.
XX Claim 10; Page 64; 748pp; English.
XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein sequence of the invention. This sequence was isolated from a HSV-2 strain SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 93. Based on homology, this sequence is a UL8 protein. The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors, activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal
SQ Sequence 381 AA;
Alignment Scores:
Pred. No.: 1.89 Length: 381
Score: 76.00 Matches: 20
Percent Similarity: 44.68% Conservative: 1
Best Local Similarity: 42.55% Mismatches: 10
Query Match: 11.01% Indels: 16
DB: 2 Gaps: 3
M64347 (1-385) x AAW72063 (1-381)
QY 365 TGGCCCCAAGA-----AGAGACACCCCTGAGCCATGGCCCTGACGCAAGCA 318
DB 239 TrpAlaProArgSerSerCysProArgGlyHisGlyArgProTrpProArgArgPro 258
QY 317 AGGCACAGCTGCCAGACTCAGG-----CCC 291
DB 259 AlaAlaSerCysProProPheGlyLysArgTrpArgGlyThrProArgProProPro 278
QY 290 AGTAAC---AGTACAGAACCA 273
||||| ||||| |||||

```

Db      279 SerAsnTyrSerThrProArg 285
RESULT 5
AAU80189
ID      AAU80189 standard; protein; 966 AA.
XX      AC
XX      AAU80189;
XX      DT
XX      15-JUL-2002 (first entry)
XX      DE
XX      Human TSP1 domain containing protein encoded by cDNA FGO1869.
XX      KW
XX      TSP1; thrombospondin domain; FGO1869; angiogenesis; vasculogenesis;
XX      KW
XX      human.
XX      OS
XX      Homo sapiens.
XX      FN
XX      JP2002085059-A.
XX      PD
XX      26-MAR-2002.
XX      PF
XX      08-SEP-2000; 2000JP-00273778.
XX      PR
XX      08-SEP-2000; 2000JP-00273778.
XX      XX
XX      (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX      PA
XX      (YOSH) YOSHITOMI PHARM IND KK.
XX      DR
XX      WPI; 2002-378268/41.
XX      DR
XX      N-PSDB; ABK50390.
XX      PT
XX      TSP1 domain-containing polypeptide useful for drug compositions.
XX      PS
XX      Claim 1; Page 30-33; 51pp; Japanese.
XX      CC
XX      The invention relates to a TSP1 (thrombospondin 1) domain-containing
XX      CC
XX      polypeptide comprising the proteins appearing as AAU80188 and AAU80189,
XX      CC
XX      encoded by cDNAs designated FGO8969 and FGO1896. Also included are
XX      CC
XX      proteins that are 50% homologous to the proteins and a polypeptide having
XX      CC
XX      at least one deletion, replacement, addition or insertion of amino acid
XX      CC
XX      in the proteins and having at least 8 repetitions of the TSP1 domain. The
XX      CC
XX      polypeptide can be used in drug compositions particularly for disorders
XX      CC
XX      associated with angiogenesis and vasculogenesis. The present sequence is
XX      CC
XX      the TSP1 domain containing protein encoded by cDNA FGO1869
XX      SQ
XX      Sequence 966 AA;

Alignment Scores:
Pred. No.:      3.3      Length:      966
Score:          75.00    Matches:      41
Percent Similarity: 29.79% Conservative: 15
Best Local Similarity: 21.81% Mismatches: 51
Query Match:      10.87% Indels:      81
DB:              5      Gaps:       9

M64347 (1-385) x AAU80189 (1-966)
QY      385 CACCTCTGGCCACATGCACTGGCCGCCAAGAGAGACACACCTGAGCCATGGCCCTGCA 326
D      460 HisArgSerProCysLeuGlyProAspThrGlnThrArgGlnGlnProCysProGly 479
QY      325 -----GGCAAGCAAGGACAGCTGCCACGACTCCAGCTCAGGCGCCA--- 290
D      480 LeuLeuGluAlaCysSerTrpGlyProTyrProCysSerArgSerCysGlyProGly 499
QY      289 -----GTACAGTACAGACCAACCACTGATTC 260
D      500 LeuAlaSerArgSerGlySerCysProCysLeuMetAlaLysAlaAspProThrCysAsn 519
QY      259 ACGGCTTCCCTCCAA-----GCTTTGAAGGTAGCAGTCCAGGCTATAAA 215
D      520 SerThrPheLeuHisLeuAspThrGlnGlyCysTyrSerGlyProCysPro----- 536

```

```

QY      214 ACTCTAGAAGCATTGCGGTA---AGAAGTGTTAAGTCTACACAAATACATCTTGTAAAAA 158
D      537 -----Glu-GluCysValTrpSerSerTrpSerTrpThrArgCysSerCysArgVa 554
QY      157 CTCATAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 98
D      554 lIeuValGlnGlnArgTyrArg----- 561
QY      97 GCACACAGGCGCGCC----- 82
D      562 -HisGlnGlyProAlaSerArgGlyAlaArgAlaGlyAlaProCysThrArgLeuAspG1 581
QY      81 -----CCTCCTCGAA 71
D      581 yHisPheArgProCysLeuIleSerAsnCysSerGluAspSerCysThrProPheG1 601
QY      70 A-----CCGTCCTCCCTGCTGGGACACACAGCAATT-----AGAAGAAT 32
D      601 upheHisAlaCysGlySerProCysAlaGlyLeuCysAlaThrHisLeuSerHisGlnLe 621
QY      31 TTGTATGAAATACACAGCTTGC 10
D      621 uCysGlnAspLeuProProCys 628

RESULT 6
ABB90106
ID      ABB90106 standard; protein; 147 AA.
XX      AC
XX      ABB90106;
XX      DT
XX      24-MAY-2002 (first entry)
XX      DE
XX      Human polypeptide SEQ ID NO 2482.
XX      KW
XX      Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
XX      KW
XX      antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
XX      KW
XX      vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX      KW
XX      cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX      OS
XX      Homo sapiens.
XX      PN
XX      WO200190304-A2.
XX      PD
XX      29-NOV-2001.
XX      PF
XX      18-MAY-2001; 2001WO-US016450.
XX      PR
XX      19-MAY-2000; 2000US-0205515P.
XX      PA
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      PI
XX      Birse CE, Rosen CA;
XX      DR
XX      WPI; 2002-122018/16.
XX      DR
XX      N-PSDB; ABL90515.
XX      PT
XX      Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
XX      PT
XX      prevention of neural, immune system, muscular, reproductive,
XX      PT
XX      gastrointestinal, pulmonary, cardiovascular, renal and proliferative
XX      disorders.
XX      PS
XX      Claim 11; SEQ ID NO 2482; 2081pp + Sequence Listing; English.
XX      CC
XX      The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX      CC
XX      (ABB90140-ABB90444) useful for preventing, treating or ameliorating
XX      CC
XX      medical conditions e.g. by protein or gene therapy. The genes are
XX      CC
XX      isolated from a range of human tissues disclosed in the specification.
XX      CC
XX      The nucleic acids, proteins, antibodies and (anti)agonists are useful in
XX      CC
XX      the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX      CC
XX      ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
XX      CC
XX      breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX      CC
XX      disorders e.g. Addison's disease, allergies, autoimmune haemolytic

```

CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 147 AA;

Alignment Scores:
 Pred. No.: 4.13 Length: 147
 Score: 72.50 Matches: 34
 Percent Similarity: 42.06% Conservative: 19
 Best Local Similarity: 26.98% Mismatches: 40
 Query Match: 10.33% Indels: 33
 DB: 5 Gaps: 8

M64347 (1-385) x ABB90106 (1-147)

QY 43 TGTGTGTCTCCAGCAGGAGACGGTTTCCAGG-----AGGGCGCGGCC----- 90
 DB 9 CysAlaSerArgGlyProSerPheProSerLeuHisGlyArgProLeuGlu 28
 QY 91 -----TGTGTGCAGGT-----CGATGTTATTAGATTACAGTTTATATATA 135
 DB 29 ProProCysThrGluLeuTrpAlaProPheCysAsnValThrLeuLysPheLeu 48
 QY 136 TCTATATATATATTTATTAGTTTTTACAGATGTTTGTGTAGACTTAACTTCT 195
 DB 49 ValLeuPhe-----LeuCysCysAsnValAsnThrGly 59
 QY 196 TAGCGAATGCTTAGAGTTTATAGCTGGAGTCTACTTCAAGCTTGAGGGAG 255
 DB 60 Phe-----LeuAspPhePheLeu-----MetGlyValLeuGlyGlyThr 73
 QY 256 CGTGAATTCAGTTGTTCTGTCTGTTACTGGGCGCTGAGTCT-----GGGCAG 309
 DB 74 AspGlyValArgGluAlaProProTrpLeuValGluGlyThrCysValAlaAla 93
 QY 310 CTGTCCCTGCTGCTGCTGAGGCG-----CATGGCTCAGGGTGG 348
 DB 94 LeuSerLeuProLeuProGlyCysLysProGluValCysAlaAlaGlyArgGlyCys 113
 QY 349 TCTCTCTTGGGCGCCAG 366
 DB 114 SerGlyLeuGlyLeuGln 119

RESULT 7

AAU80188
 ID AAU80188 standard; protein; 961 AA.

XX AC AAU80188;

XX DT 15-JUL-2002 (first entry)

XX DE Human TSPI domain containing protein encoded by cDNA FGO6969.

XX KW TSPI; thrombospondin domain; FGO6969; angiogenesis; vasculogenesis;

XX KW human.

XX OS Homo sapiens.

XX PN JF2002085059-A.

XX PD 26-MAR-2002.

XX PF 08-SEP-2000; 2000JP-00273778.

XX PR 08-SEP-2000; 2000JP-00273778.

XX PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.

PA (YOSH) YOSHITOMI PHARM IND KK.

XX WPI; 2002-378268/41.

XX N-PSDB; ABK50389.

XX TSPI domain-containing polypeptide useful for drug compositions.

XX Claim 1; Page 23-26; 51pp; Japanese.

XX The invention relates to a TSPI (thrombospondin 1) domain-containing

CC polypeptide comprising the proteins appearing as AAU80188 and AAU80189,

CC encoded by cDNAs designated FGO6969 and FGO1896. Also included are

CC proteins that are 90% homologous to the proteins and a polypeptide having

CC at least one deletion, replacement, addition or insertion of amino acid

CC in the proteins and having at least 8 repetitions of the TSPI domain. The

CC polypeptide can be used in drug compositions particularly for disorders

CC associated with angiogenesis and vasculogenesis. The present sequence is

CC the TSPI domain containing protein encoded by cDNA FGO6969

XX SQ Sequence 961 AA;

Alignment Scores:
 Pred. No.: 10.9 Length: 961
 Score: 71.00 Matches: 35
 Percent Similarity: 33.33% Conservative: 15
 Best Local Similarity: 23.33% Mismatches: 39
 Query Match: 10.29% Indels: 61
 DB: 5 Gaps: 9

M64347 (1-385) x AAU80188 (1-961)

QY 385 CACCTCTGGCCACCATCTGGSCCCCAAGACAGACACCCCTGAGCCATGGCCCTGCA 326
 DB 511 HisArgSerProProCysLeuGlyProAspThrGlnThrArgGlnGlnProCysProGly 530
 QY 325 -----GGCAAGCAAGGGACAGCTCCCAAGCTCAGCTAGGCGCCAGTA 287
 DB 531 LeuLeuGluAlaCysSerTrpGlyProTrpGlyProCysSerArgSerCysGlyPro--- 549
 QY 286 ACAGTACAGAACGACCACTGAATTCACGGCTTCCCTCCCAAGCTTTGAAAGGTAGCAGT 227
 DB 550 -----GlyLeuAlaSerArgSerGlySerCys 558
 QY 226 CCAGGCTATAAACTCTAGAAGCATTCGTAAGAAG-----TGTTAAGTCTAC 179
 DB 559 Pro-----CysLeuMetAlaLysAlaAspProThrCys----- 569
 QY 178 AACAAATACATCTTGTAATAAACTCAATAATATATATATATATATATAAAGTTGTA 119
 DB 570 -----AsnSerThrPheLeu----- 574
 QY 118 CATCTAATAACATCGGAACCTGCACACAGCGGCG-----GCCCTCC 77
 DB 575 HisLeuAspThr-GlnGlyCysThrSerGlyProCysProGluAspSerCysThrProPr 594
 QY 76 CTGGAAA-----CCGTCTCCCTCCCTGGGACACACAGCAAT-----AG 38
 DB 594 oPheGluPheHisAlaCysGlySerProCysAlaGlyLeCysAlaThrHisLeuSerHis 614
 QY 37 AAGAAATTTGTATGAATAATACACGCTTGC 10
 DB 614 sGlnLeuCysGlnAspLeuProProCys 623

RESULT 8

ABO07129

ID ABO07129 standard; protein; 1045 AA.

XX AC ABO07129;

XX DT 13-AUG-2003 (first entry)

XX DE Novel human protein NOV18a.

KW NOVX; autoimmune disease; allergy; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; multiple sclerosis; addiction;
 KW anxiety; pain; diabetes; glomerulonephritis; obesity;
 KW systemic lupus erythematosus; asthma; scleroderma; pancreatitis;
 KW graft versus host disease; ulcer; anaemia; cancer; trauma; infection;
 KW cardiomyopathy; atherosclerosis; hypertension; AIDS; Crohn's disease;
 KW acquired immunodeficiency syndrome; chromosomal mapping; tissue typing;
 KW forensic biology; predictive medicine; gene therapy; human.

XX Homo sapiens.

XX WO200298900-A2.

XX 12-DEC-2002.

XX 04-JUN-2002; 2002WO-US017558.

XX 04-JUN-2001; 2001US-0295607P.

XX 04-JUN-2001; 2001US-0295661P.

XX 06-JUN-2001; 2001US-0296404P.

XX 06-JUN-2001; 2001US-0296418P.

XX 07-JUN-2001; 2001US-0296575P.

XX 11-JUN-2001; 2001US-0297414P.

XX 12-JUN-2001; 2001US-0297567P.

XX 15-JUN-2001; 2001US-0298528P.

XX 18-JUN-2001; 2001US-0299133P.

XX 19-JUN-2001; 2001US-0299230P.

XX 21-JUN-2001; 2001US-0299949P.

XX 22-JUN-2001; 2001US-0300177P.

XX 26-JUN-2001; 2001US-0300893P.

XX 28-JUN-2001; 2001US-0301530P.

XX 28-JUN-2001; 2001US-0301550P.

XX 03-JUL-2001; 2001US-0302951P.

XX 12-SEP-2001; 2001US-0318727P.

XX 27-SEP-2001; 2001US-0325685P.

XX 22-FEB-2002; 2002US-0358814P.

XX 03-JUN-2002; 2002US-00161927.

XX (CURA-) CURAGEN CORP.

XX Zerhusen BD, Kekuda R, Spytek KA, Shenoy SG, Miller CE, Hjalte T;

PI Gerlach VL, Baumgartner JC, Guo X, Gangolli EA, Vernet CAM;

PI Padigaru M, Li L, Pena CEA, Gorman L, Anderson DW, Edinger SR;

PI Patturajan M, Stone DJ;

XX WPI; 2003-140585/13.

XX N-PSDB; ACD13203.

Novel isolated NOVX polypeptide useful treating or preventing disorders or syndromes such as autoimmune disease, allergies, Alzheimer's disease, stroke, Parkinson's disease, Huntington's disease or multiple sclerosis.

Claim 1; Page 144; 408pp; English.

The invention describes an isolated NOVX polypeptide (I) comprising a sequence selected from a sequence (S1) of 1121, 635, 299, 1720, 176, 583, 214, 395, 1098, 134, 427, 1333, 407, 806, 804, 1253, 382, 1045, 284, 496, 506, 759, 390, 133, 215, 240, 1069, 116, 439, 1138, 477, 316, 269, 219, 305, 406, 460, 365, 380, 829 or 326 amino acids fully defined in the specification, and the mature form of S1. (I) is useful for treating or preventing a pathology associated with (I) in a subject, preferably human, or for identifying an agent that binds to (I), where the agent is a cellular receptor or a downstream effector. (I), a polynucleotide (II) encoding (I) or an anti-(I)-antibody (V) is useful treating or preventing disorders or syndromes such as autoimmune disease, allergies, Alzheimer's disease, stroke, Parkinson's disease, Huntington's disease, multiple sclerosis, addiction, anxiety, pain, diabetes, glomerulonephritis, systemic lupus erythematosus, asthma, scleroderma, graft versus host disease, pancreatitis, obesity, ulcers, anaemia, cancer, trauma, viral, bacterial or parasitic infections, cardiomyopathy, atherosclerosis, hypertension, acquired immunodeficiency syndrome (AIDS) or Crohn's disease. (I), (II) or (V) is useful in screening assays, detection assays (e.g., chromosomal mapping, tissue typing, forensic biology), predictive

CC medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials and pharmacogenomic), and in methods of treatment (e.g., therapeutic and prophylactic). (II) is useful in gene therapy, to express (I), to detect NOVX mRNA or a genetic lesion in a NOVX gene, and to modulate NOVX activity. This is the amino acid sequence of a novel human NOV protein

XX Sequence 1045 AA;

Alignment Scores:

Pred. No.:	11.1	Length:	1045
Score:	71.50	Matches:	24
Percent Similarity:	40.24%	Conservative:	9
Best local Similarity:	29.27%	Mismatches:	25
Query Match:	10.11%	Indels:	24
DB:	6	Gaps:	5

M64347 (1-385) X ABO07129 (1-1045)

QY 189 CACTTCTTACCAATGCTTCTAGAGTTTATAGCTGCTGCTGCTACCT-----TTCAAA 242

Db 225 HisArgLeuHisAsnSerValSerAlaProSerProGlyIleProProArgAspPheLys 244

QY 243 GCT-----TGGAGGGAAGCGGTGAATCA 266

Db 245 SerLeuAlaLeuAlaArgAlaProGlyHisGlyGlyPheTrp----GinGlyValAlaAla 263

QY 267 GTTGCTTCGTTCTCTACTGTTACTGGG-----CCCTGAGTCTGGCGAGCTGTC 314

Db 264 GluGlyValGlyCysThrLeuThrGlyAlaTrpArgSerProValProTrpSerGlyThr 283

QY 315 CCTGCTTCGCTGCTGCA-----GGGCCATGGCTCAGGGTGGTCTCTCTCTGG 359

Db 284 GlyCysValProGlyGlyPheThrValProGlyProArgProProAlaProAlaProTrp 303

QY 360 GGCCCA 365

Db 304 GlyPro 305

RESULT 9

ADCS1662

ID ADCS1662 standard; protein; 2273 AA.

AC ADCS1662;

DT 18-DEC-2003 (first entry)

DE Human MEGF8 protein #2.

KW Human; membrane binding-MEGF8; secretory-MEGF8; plexin domain;

KW axially fibre induction; nerve cell; heart development;

KW skeleton development; immune response; vasculogenesis; cancer growth;

KW metastasis; cytostatic; neuroprotective; cardiant; immunomodulatory;

XX Homo sapiens.

OS JP2002360254-A.

PN 17-DEC-2002.

PF 21-MAY-2001; 2001JP-00151059.

PR 27-MAR-2001; 2001JP-00090438.

XX (KAZU-) ZH KAZUSA DNA KENKYUSHO.

XX WPI; 2003-516152/49.

XX N-PSDB; ADCS1661.

XX A membrane binding- or secretory-MEGF8 gene and a protein encoded by the gene, for diagnosis and treatment of diseases caused by plexin domain.

PS Claim 1; SEQ ID NO 4; 53pp; Japanese.

XX The present invention relates to novel human membrane binding- or
 CC secretory-MEGF8 proteins, and the polynucleotide sequence encoding them.
 CC The sequences are useful in the identification of a plexin domain for the
 CC diagnosis and treatment of diseases caused by plexin domains
 CC participating in the induction of axial fibres of nerve cells,
 CC development of the heart and skeleton, immune response, vasculogenesis,
 CC and the growth and metastasis of cancer. The present represents a human
 CC membrane binding- or secretory-MEGF8 protein.

SQ Sequence 2273 AA;

Alignment Scores:
 Pred. No.: 13.8 Length: 2273
 Score: 71.00 Matches: 25
 Percent Similarity: 43.06% Conservative: 6
 Best Local Similarity: 34.72% Mismatches: 25
 Query Match: 10.11% Indels: 16
 DB: 7 Gaps: 3

M64347 (1-385) x ADCS1662 (1-2273)

QY 182 GACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTTCAA 241
 DB 752 GluLeuThrLeuLeu-----TrpAspArgThr----- 760
 QY 242 AGCTTGGAGGGAAGCGGTAATTCAAGTTGGTTGGTTCTGTACTGTACTGGGCC----- 295
 DB 761 GlyValProGlyGlySerGluIleSerPhePheLeuGluProTyrArgSerSerSer 780
 QY 296 ---CTGAGTCTGGGACGCTGCTCCCTGCTGCTGCGAGGCCATGCTCAGGCTGCTC 352
 DB 781 CysThrSerTyrSerSerCys--LeuGlyCysLeuAlaAspGlnGlyCysGlyTyrCysL 800
 QY 353 TTCTTGGGGCCCGCATGTCATGGTGGCCAGAGGT 384
 DB 800 eutThrSerAlaThrCysHisLeuArgGlnGly 810

RESULT 11

ADD04782
 ID ADCS1660 standard; protein; 2778 AA.

XX ADCS1660;

XX DT 18-DEC-2003 (first entry)

XX Human MEGF8 protein #1.

XX KW Human; membrane binding-MEGF8; secretory-MEGF8; plexin domain;

XX KW axial fibre induction; nerve cell; heart development;

XX KW skeleton development; immune response; vasculogenesis; cancer growth;

XX KW metastasis; cytostatic; neuroprotective; cardiac; immunomodulatory;

XX KW vulnery.

XX OS Homo sapiens.

XX PN JP2002360254-A.

XX PD 17-DEC-2002.

XX PF 21-MAY-2001; 2001JP-00151059.

XX PR 27-MAR-2001; 2001JP-00090438.

XX PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.

XX DR WPI; 2003-516152/49.

XX DR N-PSDB; ADCS1659.

XX PT A membrane binding- or secretory-MEGF8 gene and a protein encoded by the
 PT gene, for diagnosis and treatment of diseases caused by plexin domain.

XX

PS Claim 1; SEQ ID NO 2; 53pp; Japanese.

XX The present invention relates to novel human membrane binding- or
 CC secretory-MEGF8 proteins, and the polynucleotide sequence encoding them.
 CC The sequences are useful in the identification of a plexin domain for the
 CC diagnosis and treatment of diseases caused by plexin domains
 CC participating in the induction of axial fibres of nerve cells,
 CC development of the heart and skeleton, immune response, vasculogenesis,
 CC and the growth and metastasis of cancer. The present represents a human
 CC membrane binding- or secretory-MEGF8 protein.

SQ Sequence 2778 AA;

Alignment Scores:
 Pred. No.: 14.6 Length: 2778
 Score: 71.00 Matches: 25
 Percent Similarity: 43.06% Conservative: 6
 Best Local Similarity: 34.72% Mismatches: 25
 Query Match: 10.11% Indels: 16
 DB: 7 Gaps: 3

M64347 (1-385) x ADCS1660 (1-2778)

QY 182 GACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTTCAA 241
 DB 752 GluLeuThrLeuLeu-----TrpAspArgThr----- 760
 QY 242 AGCTTGGAGGGAAGCGGTAATTCAAGTTGGTTGGTTCTGTACTGTACTGGGCC----- 295
 DB 761 GlyValProGlyGlySerGluIleSerPhePheLeuGluProTyrArgSerSerSer 780
 QY 296 ---CTGAGTCTGGGACGCTGCTCCCTGCTGCTGCGAGGCCATGCTCAGGCTGCTC 352
 DB 781 CysThrSerTyrSerSerCys--LeuGlyCysLeuAlaAspGlnGlyCysGlyTyrCysL 800
 QY 353 TTCTTGGGGCCCGCATGTCATGGTGGCCAGAGGT 384
 DB 800 eutThrSerAlaThrCysHisLeuArgGlnGly 810

RESULT 11

ADD04782
 ID ADD04782 standard; protein; 1043 AA.

XX AC ADD04782;

XX DT 01-JAN-2004 (first entry)

XX DE Pig RAG-1 amino acid sequence #SEQ ID 3.

XX KW Pig; RAG-1; recombination activation gene; human transplantation.

XX OS Sus scrofa.

XX PN WO2003066855-A1.

XX PD 14-AUG-2003.

XX PF 04-FEB-2003; 2003WO-JP001114.

XX PR 04-FEB-2002; 2002JP-00027450.

XX PA (NAAAG-) NAT INST AGROBIOLOGICAL SCI.

XX PI Yasue H, Watanabe S, Honma D, Takagaki Y, Hatsuse H;

XX DR WPI; 2003-618546/58.

XX DR N-PSDB; ADD04784.

XX PT New pig recombination activation gene 1 (RAG-1) gene and protein for
 PT production of RAG-1 knockout pigs suitable as human transplant tissue
 PT donors.

XX Claim 1; Page 46-54; 58pp; Japanese.

m64347.rag

Wed Sep 22 14:52:11 2004

PN	EPI033405-A2.	PR	01-JUL-1999;	99US-0142154P.
XX		PR	02-JUL-1999;	99US-0142055P.
PD		PR	06-JUL-1999;	99US-0142380P.
XX	06-SEP-2000.	PR	08-JUL-1999;	99US-0142803P.
PF		PR	09-JUL-1999;	99US-0142920P.
XX	25-FEB-2000; 2000EP-00301439.	PR	12-JUL-1999;	99US-0142977P.
XX		PR	13-JUL-1999;	99US-0143542P.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624P.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005P.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085P.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086P.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332P.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333P.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334P.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335P.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352P.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632P.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884P.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086P.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086P.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087P.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089P.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192P.
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388P.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389P.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038P.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204P.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302P.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192P.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260P.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303P.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416P.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0148368P.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066P.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151080P.
PR	21-JUN-1999;	PR	30-AUG-1999;	99US-0151303P.
PR	22-JUN-1999;	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	PR	01-SEP-1999;	99US-0151930P.
PR	23-JUN-1999;	PR	07-SEP-1999;	99US-0152363P.
PR	24-JUN-1999;	PR	10-SEP-1999;	99US-0153070P.
PR	28-JUN-1999;	PR	13-SEP-1999;	99US-0153758P.
PR	29-JUN-1999;	PR	15-SEP-1999;	99US-0154018P.
PR	30-JUN-1999;	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	PR		

[illegible]

```
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140363P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144894P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149823P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156595P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 15.9
Score: 70.50
Percent Similarity: 38.53%
Best Local Similarity: 27.52%
Query Match: 10.22%
DB: 3

M64347 (1-385) x AAG50615 (1-2204)

Qy 364 GGGCCCCAGAGAGACCCCTGAGC-----CATGGCCCTGCAGC-----AAG 320
Db 1467 GlyProLeuHisGlnThrProIleSerAspGlnTy-LysProLeuGlyTyLeuAspArg 1486
```

QY 319 CAGGGACAGCTGCCAGACTCAGGGCCAGTACAGTACAGAACGACCACTGAATTC 260
Db 1487 GlnArgLeuAlaAlaArgSerAsnThrThrTyrcysTyraSpPheProLeuAlaPhe 1506
QY 259 ACGGCTTCCTCCCAAGCTTTGAAAGTAGCAGT---CCAGGCTATAAAACTCTAGAAGCA 203
Db 1507 GlyThrAlaLeuGluLeuLeuTrpAlaSerGlnHisProGlyVallys----- 1522
QY 202 TTGCGTAAGAAGTGTAAAGTCTACAAATAACATCTTGTAAAACTCAATAAATTATAT 143
Db 1523 -----LysProTyr 1525
QY 142 ATATAGATATATAAATTTGTAACATCTATAACATCGGAACCTGCACACAGGCGCGGC 83
Db 1526 LysAspThrLeuIleAsnVallysGluLeuValPheSerLysProGluGlySerSerGly 1545
QY 82 CCCTCCCTCGAAACCGCTCCCTGCCT 56
Db 1546 ThrSerLeuAspLeuValGluArgPro 1554

RESULT 15
AAR76949
ID AAR76949 standard; protein; 2254 AA.
XX
AC AAR76949;
DT 28-FEB-1996 (first entry)
XX
DE ACCase.
XX
KW Polymerase chain reaction; PCR; primer; amplify; acetyl CoA carboxylase;
KW ACCase; transgenic plant; regulation; fat; protein; agricultural plant.
XX
OS Arabidopsis thaliana.
XX
FN JP07143887-A.
XX
PD 06-JUN-1995.
XX
PF 28-JUN-1994; 94JP-00146827.
XX
PR 17-AUG-1993; 93JP-00203477.
XX
PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.
XX
XX WPI; 1995-236465/31.
DR N-PSDB; AAQ93232.
XX
XX Plant acetyl CoA carboxylase gene obtd. by PCR amplification - useful for
PT increasing fat/protein content in plants.
XX
PS Claim 11; Page 14-22; 33pp; Japanese.

CC This sequence is encoded by the cDNA of the acetyl CoA carboxylase
CC (ACCcase) gene which was amplified using the primers given in AAQ93221- 30
CC and AAQ93233-36. This full length ACCase gene may be used to produce a
CC transgenic plant such that it expresses a changed amount of ACCase. The
CC ACCase gene may be used to regulate the production of fat/protein in
CC agricultural plants
XX
SQ Sequence 2254 AA;

Alignment Scores:	16	Length:	2254
Pred. NO.:	70.50	Matches:	30
Score:	38.53%	Conservative:	12
Percent Similarity:	27.52%	Mismatches:	40
Best Local Similarity:	10.22%	Indels:	27
Query Match:	2	Gaps:	4
DB:			

M64347 (1-385) x AAR76949 (1-2254)

QY 364 GGGGCCCCAAGAAGAGACCACCTGAGC-----CATGGCCCTGCAGGC-----AAG 320

Db 1482 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1501
QY 319 CAAGGGACAGCTGCCAGACTCAGGGCCAGTAAACAGTACAGAACGACCACTGAATTC 260
Db 1502 GlnArgLeuAlaAlaArgSerAsnThrThrTyrcysTyraSpPheProLeuAlaPhe 1521
QY 259 ACGGCTTCCTCCCAAGCTTTGAAAGTAGCAGT---CCAGGCTATAAAACTCTAGAAGCA 203
Db 1522 GlyThrAlaLeuGluLeuLeuTrpAlaSerGlnHisProGlyVallys----- 1537
QY 202 TTGCGTAAGAAGTGTAAAGTCTACAAATAACATCTTGTAAAACTCAATAAATTATAT 143
Db 1523 -----LysProTyr 1540
QY 142 ATATAGATATATAAATTTGTAACATCTATAACATCGGAACCTGCACACAGGCGCGGC 83
Db 1541 LysAspThrLeuIleAsnVallysGluLeuValPheSerLysProGluGlySerSerGly 1560
QY 82 CCCTCCCTCGAAACCGCTCCCTGCCT 56
Db 1561 ThrSerLeuAspLeuValGluArgPro 1569

Search completed: September 22, 2004, 10:50:46
Job time : 96.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:47:46 ; Search time 23.5 Seconds
(without alignments)
1691.574 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAGCAGCTGTA.....GTGCTGTGGCCAGAGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DRV=xlp
-Q/cgn2_1/USFTO_spool_p/CANELLA305B/runat_22092004.112337.14559/app_query.fasta_1.583
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -CUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CANELLA305B @CGN 1.1.46 @runat 22092004.112337.14559 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PTCUTS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	10.7	539	US-09-252-991A-17070	Sequence 17070, A
2	70.5	10.2	254	US-08-677-010-3	Sequence 3, Appli
3	70.5	10.2	254	US-08-790-519-3	Sequence 3, Appli
4	70	10.0	205	US-09-252-991A-16946	Sequence 16946, A
5	68.5	9.8	1040	US-08-254-989-2	Sequence 2, Appli
6	67.5	9.6	527	US-09-659-166-2	Sequence 2, Appli
7	67	9.5	528	US-08-802-466-2	Sequence 2, Appli
8	67	9.5	528	US-09-350-484-2	Sequence 2, Appli
9	67	9.7	930	US-09-122-126B-15	Sequence 15, Appl
10	67	9.7	930	US-09-634-286A-15	Sequence 15, Appl
11	66.5	9.6	2523	US-08-185-432-18	Sequence 18, Appl
12	66.5	9.6	2523	US-08-899-232-3	Sequence 3, Appli

C 13	65.5	9.5	140	4	US-09-252-991A-30876	Sequence 30876, A
C 14	65	9.4	572	4	US-09-252-991A-21105	Sequence 21105, A
C 15	65	9.4	930	4	US-09-369-364A-2	Sequence 2, Appli
C 16	64.5	9.3	160	4	US-09-252-991A-31945	Sequence 31945, A
C 17	64	9.3	154	4	US-08-489-039A-9332	Sequence 9332, Ap
C 18	64	9.3	166	4	US-09-252-991A-21902	Sequence 21902, A
C 19	64	9.3	305	4	US-10-162-012-23	Sequence 23, Appl
C 20	64	9.3	904	4	US-09-976-594-615	Sequence 615, App
C 21	64	9.1	1214	2	US-08-231-193A-54	Sequence 54, Appl
C 22	64	9.1	1214	2	US-08-486-273A-54	Sequence 54, Appl
C 23	64	9.1	1214	3	US-08-480-474-54	Sequence 54, Appl
C 24	64	9.1	1214	3	US-08-940-086A-54	Sequence 54, Appl
C 25	64	9.1	1214	4	US-08-940-035A-54	Sequence 54, Appl
C 26	64	9.1	1214	4	US-08-935-105A-54	Sequence 54, Appl
C 27	64	9.1	1214	4	US-09-648-797-54	Sequence 54, Appl
C 28	64	9.1	1214	4	US-09-386-123-54	Sequence 54, Appl
C 29	64	9.1	1219	2	US-08-231-193A-50	Sequence 50, Appl
C 30	64	9.1	1219	2	US-08-486-273A-50	Sequence 50, Appl
C 31	64	9.1	1219	3	US-08-480-474-50	Sequence 50, Appl
C 32	64	9.1	1219	3	US-08-940-086A-50	Sequence 50, Appl
C 33	64	9.1	1219	4	US-08-940-035A-50	Sequence 50, Appl
C 34	64	9.1	1219	4	US-08-935-105A-50	Sequence 50, Appl
C 35	64	9.1	1219	4	US-09-386-123-50	Sequence 50, Appl
C 36	64	9.1	1219	4	US-09-648-797-50	Sequence 50, Appl
C 37	64	9.1	1231	2	US-08-231-193A-48	Sequence 48, Appl
C 38	64	9.1	1231	2	US-08-486-273A-48	Sequence 48, Appl
C 39	64	9.1	1231	3	US-08-480-474-48	Sequence 48, Appl
C 40	64	9.1	1231	3	US-08-940-086A-48	Sequence 48, Appl
C 41	64	9.1	1231	4	US-08-940-035A-48	Sequence 48, Appl
C 42	64	9.1	1231	4	US-08-935-105A-48	Sequence 48, Appl
C 43	64	9.1	1231	4	US-09-648-797-48	Sequence 48, Appl
C 44	64	9.1	1231	4	US-09-386-123-48	Sequence 48, Appl
C 45	64	9.1	1235	2	US-08-231-193A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-09-252-991A-17070
; Sequence 17070, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17070
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17070

Alignment Scores:	0.861	Length:	539
Pred. No.:	75.00	Matches:	25
Score:	44.44%	Conservative:	3
Best Local Similarity:	39.68%	Mismatches:	27
Query Match:	10.68%	Indels:	8
DB:	4	Gaps:	3

M64347 (1-385) x US-09-252-991A-17070 (1-539)

Qy 201 AATGCTTCTAGAGTTTATAGCTGACTGCTACCTTTCAAGCTTGGAGGAGCGGTG 260

Db 29 SerAlaAlaArgSerLeuArgSerGlyLeuCysProGlyLysAlaTrpArgProAlaAla 48

QY 261 AATCAGTGGTTCCTCTCTACTGTACTGGGCGCTGAGTCTGGCAGCGTCTCCCTTGC 320
 Db 49 TrpSerValAlaSer---SerArgGlyThrGlyPro-----AlaAlaCys 62
 QY 321 TTGCTCGCAGGCGCATGGCTCAGGCTGGTCTCTCTCTTGG-----GGCCAGTGCATGCT 374
 Db 63 ArgProGlyArgProGlyArgSerValAlaArgTrpSerAlaSerProGlySerPro 82
 QY 375 GGCAGAGG 383
 Db 83 GlyArgArg 85

RESULT 2
 US-08-677-010-3
 ; Sequence 3, Application US/08677010
 ; Patent No. 5925805
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohlrogge, John B.
 ; APPLICANT: Roessler, Keith R.
 ; APPLICANT: Shorrock, Basil S.
 ; TITLE OF INVENTION: Structure and Expression of an
 ; TITLE OF INVENTION: Arabidopsis Acetyl-coenzyme A Carboxylase Gene
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: Michigan
 ; COUNTRY: U.S.A.
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/677,010
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 8550-00002CPA
 ; TELEPHONE: (810)641-1600
 ; TELEFAX: (810)641-0270
 ; INFORMATION FOR SEQ ID NO: 3:
 ; LENGTH: 2254 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-677-010-3

Alignment Scores:
 Pred. No.: 4.65 Length: 2254
 Score: 70.50 Matches: 30
 Percent Similarity: 38.53% Conservative: 12
 Best Local Similarity: 27.52% Mismatches: 40
 Query Match: 10.22% Indels: 27
 DB: 2 Gaps: 4

M64347 (1-385) x US-08-677-010-3 (1-2254)

QY 364 GGGCCCCAAGAGAGACACCTGAGC-----CATGGCCTGCAGGC-----AAG 320
 Db 1482 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1501
 QY 319 CAAGGACAGCTGCCAGACTCAGGCCCGCCAGCTACAGACGACCACTGAATTC 260
 Db 1502 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1521

QY 259 ACGGCTCCCTCCCAAGCTTTGAAGAGTAGCAGT---CCAGGCTATAAACTCTAGAAGCA 203
 Db 1522 GlyThrAlaLeuGluLeuLeuTrpAlaSerGlnHisProGlyValLys-----1537
 QY 202 TTGCGTAAGAAGTGTTAAGTCTACACAATAATACATCTTGTAAATAACTCAATAATATAT 143
 Db 1538 -----LysProTyr 1540

QY 142 ATATAGATATATATAAACTTTGTAACATCTAATACATCGGAACCTGCACACAGGCGCGC 83
 Db 1541 LysAspThrLeuLeuAsnValLysGluLeuValPheSerLysProGluGlySerSerGly 1560

QY 82 CCTCCCTCGGAACCGTCTCCCTGCCT 56
 Db 1561 ThrSerLeuAspLeuValGluArgPro 1569

RESULT 3
 US-08-790-519-3
 ; Sequence 3, Application US/08790519
 ; Patent No. 5962767
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohlrogge, John B.
 ; APPLICANT: Roessler, Keith R.
 ; APPLICANT: Shorrock, Basil S.
 ; TITLE OF INVENTION: Structure and Expression of an
 ; TITLE OF INVENTION: Arabidopsis Acetyl-coenzyme A Carboxylase Gene
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce
 ; STREET: P.O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: Michigan
 ; COUNTRY: U.S.A.
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/790,519
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/248,630
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.
 ; REGISTRATION NUMBER: 36683
 ; REFERENCE/DOCKET NUMBER: 6550-00002
 ; TELEPHONE: (810)641-1600
 ; TELEFAX: (810)641-0270
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2254 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-790-519-3

Alignment Scores:
 Pred. No.: 4.65 Length: 2254
 Score: 70.50 Matches: 30
 Percent Similarity: 38.53% Conservative: 12
 Best Local Similarity: 27.52% Mismatches: 40
 Query Match: 10.22% Indels: 27
 DB: 2 Gaps: 4

M64347 (1-385) x US-08-790-519-3 (1-2254)

QY 364 GGGCCCCAAGAGAGACACCTGAGC-----CATGGCCTGCAGGC-----AAG 320

```
Db 1482 GlyProLeuHisGluThrProIleSerAspGlnTyrLysProLeuGlyTyrLeuAspArg 1501
QY 319 CAAGGGAGACGTGCCAGACTCAGGCGCCAGTAAACAGTACAGAACGAAACCAACTGAATTC 260
Db 1502 GlnArgLeuAlaAlaArgSerAsnThrThrTyrCysTyrAspPheProLeuAlaPhe 1521
QY 259 ACGCTTCCTCCCAAGCTTTGAAGGTAGAGT---CCAGGCTATAAACTCTAGAAGCA 203
Db 1522 GlyThrAlaLeuGluLeuLeuTrpAlaSerGlnHisProGlyValLys----- 1537
QY 202 TTGGTGAAGAAGTGTAAAGTCTACAAACAATACATCTGTAAACAACTCAATAAATATAT 143
Db 1538 -----LysProTyr 1540
QY 142 ATATAGATATATATAACTTGTAACTCTAATAACATCGGAACCTCCACACAGGCGCGC 83
Db 1541 LysAspThrLeuIleAsnValLysGluLeuValPheSerLysProGluGlySerGly 1560
QY 82 CCCTCCCTGGAACCGTCTCCCTGCCT 56
Db 1561 ThrSerLeuAspLeuValGluArgPro 1569

RESULT 4
US-09-252-991A-16946
; Sequence 16946, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16946
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16946

Alignment Scores:
Pred. No.: 2.76 Length: 205
Score: 70.00 Matches: 31
Percent Similarity: 31.62% Conservative: 12
Best Local Similarity: 22.79% Mismatches: 32
Query Match: 9.97% Indels: 61
DB: 4 Gaps: 7

M64347 (1-385) x US-09-252-991A-16946 (1-205)
QY 1 GACTTCAAGCAAGCTGGTATTTCATACAATCTCTCT-----AATGCTGTGTGTC 54
Db 30 AsnTrpArgAlaAlaTrpIlePheProProSerSerThrAlaArgCysCys----- 47
QY 55 CAGGCGAGGAGACGGTTTCCAGGAGGCGCGCCCTGTGTGCAGTTCGAGTGTATTA 114
Db 48 ---AlaValArgArgTrpAlaAlaSerTrpProAlaSerLys---Pro-Cys----- 63
QY 115 GATGTTACAAGTTTATATATATATATATATATATATATATATATATATATATATAT 174
Db 63 ----- 63
QY 175 TGTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTAC 234
Db 64 -----TrpProAla 67
QY 235 CTTTCAAGCTGGAGGGAAGCGGTGAATTCAGTTGGTTTGGTCTGTACTGTACTGGGC 294
|||||:|||||
|||||:|||||
```

```
Db 67 rValSerProAlaAsnGlyAlaAlaArgSerAlaCysTrpIleSerLeuAlaAlaCysTrpAr 87
QY 295 CCTGAGTCTGGCAGCTGT-----CCCTTGCTT-----GCCTG 327
Db 87 gArgAlaAlaProAlaCysArgGlyProIleLeuGlyAlaArgLeuSerProAlaCys 107
QY 328 CAGGCGCAATGGCTCAGGCTGTCTCTTCTTGGGGCCCGCAGTGCATGG 373
Db 107 gArgArgCysSer-----ProSerAlaTrp 115

RESULT 5
US-08-254-989-2
; Sequence 2, Application US/08254989
; Patent No. 5859307
; GENERAL INFORMATION:
; APPLICANT: Mombaerts, Peter
; APPLICANT: Tonedawa, Susumu
; APPLICANT: Johnson, Randall S.
; APPLICANT: Papalocannou, Virginia
; TITLE OF INVENTION: Mutant RAG-1 Deficient Animals Having No
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick & Cody
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,989
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/830831
; FILING DATE: 04-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabat, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MITS783
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-8555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1040 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mouse
US-08-254-989-2

Alignment Scores:
Pred. No.: 6.66 Length: 1040
Score: 68.50 Matches: 23
Percent Similarity: 44.59% Conservative: 10
Best Local Similarity: 31.08% Mismatches: 22
Query Match: 9.76% Indels: 19
DB: 2 Gaps: 4

M64347 (1-385) x US-08-254-989-2 (1-1040)
QY 84 CCGGCCCTGTGTGCGAGTTCGAGTTATTAGATTGTTACAAGTTTATATATATCTATATA 143
|||||:|||||
|||||:|||||
```



```

Alignment Scores:
Pred. No.: 9.93 Length: 930
Score: 67.00 Matches: 38
Percent Similarity: 30.22% Conservative: 17
Best Local Similarity: 20.88% Mismatches: 54
Query Match: 9.71% Indels: 73
DB: 4 Gaps: 7

M64347 (1-385) x US-09-634-286A-15 (1-930)
QY 379 TGGCCACATGCACT----- 365
DB 452 TrpSerLysCysThrSerAlaThrIleThrGluPheLeuAspGlyHisGlyAsnCys 471
QY 364 -----GGGCCCCAAGAA----- 353
DB 472 LeuLeuAspLeuProArgLysGlnIleLeuGlyProGluGluLeuProGlyGlnThrTyr 491
QY 352 -----GAGACCACCTGAGCCATGGCCCT-----GCAGCAAGCAA 317
DB 492 AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrSerValCysProGlyMet 511
QY 316 GGGACAGCTGCCAGACTCAGGGCCCACTAAGTACAGACGAAACCACTGAATTCACG 257
DB 512 AspValCysAlaArgLeuTyrCysAlaValValArgGlnGlyGlnMetValCysLeuThr 531
QY 256 GTTCCCTCCCAAGCTTTGAAGGTAGCAGTCCAGGCTAT----- 218
DB 532 LysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgIleCysLeuGlnGly 551
QY 217 AAAAAGCTTAGAAGCAATTCGTAAGTCTTACAAACA----- 175
DB 552 LysCysValAspLysThrLysLysLys-TyrTyrSerThrSerHisGlyAsnTrpG1 571
QY 174 -----AATCATCTTGTAAAACTCAATAAATATATATA 140
DB 571 yserTrpGlySerTrpGlyGlnCysSerArgSerCysGlyGlyValGlnPheAlaTy 591
QY 139 TAGATATATATAAACTTGTAACTCTAATAACATCGGAACCTGCACACAGGGCGGCCCC 80
DB 591 rArgHis-----CysAsn-----AsnProAlaPr 599
QY 79 TCCCTGGAACCGTCTCCCTCGCTGGACACACAGCAATTAGAAGATTGTATGAAAT 20
DB 599 cArgAsnAsnGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSerLeuMe 619
QY 19 ACCA 16
DB 619 tPro 620

RESULT 11
US-08-185-432-18
; Sequence 18, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTAX PROTEINS, NUCLEIC ACIDS, AND
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

Alignment Scores:
Pred. No.: 15.1 Length: 2523
Score: 66.50 Matches: 20
Percent Similarity: 39.06% Conservative: 5
Best Local Similarity: 31.25% Mismatches: 26
Query Match: 9.64% Indels: 13
DB: 1 Gaps: 3

M64347 (1-385) x US-08-185-432-18 (1-2523)
QY 201 TGCCTAAGCAAGTGTAGTCTTACAAACAATACATCTTGTAAAACTCA-----ATA 151
DB 176 CysLysGlnAspIleAsnGluCysSerGlnAsnProCysLysAsnGlyGlnCysIle 195
QY 150 AATATATATATAGATATATATATAAACTGTAACTCTTAATTAACATCGGAACCTGCACACA 91
DB 196 AsnGluPheGlySerTyrArgCysThrCys-----GlnAsnArgPheThr 210
QY 90 GGG-----CCGGCCCTCCCTCGTAAACCGCTCTCCCTCGTGGACACACA 46
DB 211 GlyArgAsnCysAspGluProTyrValProCysAsnProSerProCysLeuAsnGlyGly 230
QY 45 GCAATTAGAAGA 34
DB 231 ThrCysArgGln 234

RESULT 12
US-08-899-232-3
; Sequence 3, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Oii, HuiLin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenopus sp.
US-08-899-232-3

Alignment Scores:
Pred. No.: 15.1 Length: 2523
Score: 66.50 Matches: 20
Percent Similarity: 39.06% Conservative: 5
Best Local Similarity: 31.25% Mismatches: 26
Query Match: 9.64% Indels: 13
DB: 1 Gaps: 3

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2523 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-18

Alignment Scores:
Pred. No.: 9.93 Length: 930
Score: 67.00 Matches: 38
Percent Similarity: 30.22% Conservative: 17
Best Local Similarity: 20.88% Mismatches: 54
Query Match: 9.71% Indels: 73
DB: 4 Gaps: 7

M64347 (1-385) x US-09-634-286A-15 (1-930)
QY 379 TGGCCACATGCACT----- 365
DB 452 TrpSerLysCysThrSerAlaThrIleThrGluPheLeuAspGlyHisGlyAsnCys 471
QY 364 -----GGGCCCCAAGAA----- 353
DB 472 LeuLeuAspLeuProArgLysGlnIleLeuGlyProGluGluLeuProGlyGlnThrTyr 491
QY 352 -----GAGACCACCTGAGCCATGGCCCT-----GCAGCAAGCAA 317
DB 492 AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrSerValCysProGlyMet 511
QY 316 GGGACAGCTGCCAGACTCAGGGCCCACTAAGTACAGACGAAACCACTGAATTCACG 257
DB 512 AspValCysAlaArgLeuTyrCysAlaValValArgGlnGlyGlnMetValCysLeuThr 531
QY 256 GTTCCCTCCCAAGCTTTGAAGGTAGCAGTCCAGGCTAT----- 218
DB 532 LysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgIleCysLeuGlnGly 551
QY 217 AAAAAGCTTAGAAGCAATTCGTAAGTCTTACAAACA----- 175
DB 552 LysCysValAspLysThrLysLysLys-TyrTyrSerThrSerHisGlyAsnTrpG1 571
QY 174 -----AATCATCTTGTAAAACTCAATAAATATATATA 140
DB 571 yserTrpGlySerTrpGlyGlnCysSerArgSerCysGlyGlyValGlnPheAlaTy 591
QY 139 TAGATATATATAAACTTGTAACTCTAATAACATCGGAACCTGCACACAGGGCGGCCCC 80
DB 591 rArgHis-----CysAsn-----AsnProAlaPr 599
QY 79 TCCCTGGAACCGTCTCCCTCGCTGGACACACAGCAATTAGAAGATTGTATGAAAT 20
DB 599 cArgAsnAsnGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSerLeuMe 619
QY 19 ACCA 16
DB 619 tPro 620

RESULT 11
US-08-185-432-18
; Sequence 18, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTAX PROTEINS, NUCLEIC ACIDS, AND
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

	Gaps:	3
M64347 (1-385) x US-08-999-232-3 (1-2523)		
Qy	201 TGCCTAAGAAAGTGTTAAGTCATCAACAATAACATCCTTGTA AAAACTCA-----ATA 151 : : : :	
Db	176 CyslysGlnAsnPlieAsnGluCySerGlnAsnProCysLysAsnGlyGlnCysIle 195 : : : :	
Qy	150 AATTATATATATAGATATATATAAACCTTTGGTAACATCTATAAACATCGGAACCTGCCACACA 91 	
Db	196 AsnGluPheGlySerTyArgCysThrCys-----GlnAsnArgPheThr 210 	
Qy	90 GGG-----CCGGCCCCCTCCCTGGAAAACCGTCTCCCTGCCTGGGACACACA 46 	
Db	211 GlyArgAsnCysaspGluProTyValProCysAsnProserProcysLeuAsnGlyGly 230 	
Qy	45 GC AAATTAGAACA 34 :	
Db	231 ThrCysArgGln 234 :	
RESULT 13		
US-09-252-991A-30876		
; Sequence 30876, Application US/09252991A		
; Patent No. 6551795		
; GENERAL INFORMATION:		
; APPLICANT: Marc J. Rubenfield et al.		
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS		
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS		
; FILE REFERENCE: 107196.136		
; CURRENT APPLICATION NUMBER: US/09/252,991A		
; CURRENT FILING DATE: 1999-02-18		
; PRIOR APPLICATION NUMBER: US 60/074,788		
; PRIOR FILING DATE: 1998-02-18		
; PRIOR APPLICATION NUMBER: US 60/094,190		
; PRIOR FILING DATE: 1998-07-27		
; NUMBER OF SEQ ID NOS: 33142		
; SEQ ID NO 30876		
; LENGTH: 140		
; TYPE: PRF		
; ORGANISM: Pseudomonas aeruginosa		
US-09-252-991A-30876		
Alignment Scores:		
Pred. No.: 9.03 Length: 140		
Score: 65.50 Matches: 17		
Percent Similarity: 44.90% Conservative: 5		
Best Local Similarity: 34.69% Mismatches: 14		
Query Match: 9.49% Indels: 13		
: 4 Gaps: 3		
M64347 (1-385) x US-09-252-991A-30876 (1-140)		
Qy	344 CCTGAGCGCATGGCCCTCCAGCAAGCA-----AGGGACAGCTGCCCA 303 	
Db	80 ProSerProTrpProCysArgProAlaProValSerArgArgTyArgProAlaMetPro 99 	
Qy	302 GAC-----TCAGGGCCCAGTACAGTACAGACAGAACCAACTGAATTACGGCTTCCCT 249 	
Db	100 SerSerProSerAlaProproThrSetSerArgCysSer-----Pro 114 	
Qy	248 CCAAGCTTTGAAGGTAGCAGTCCAGG 222 	
Db	115 ProGlyTrpargArgGlySerAlaarg 123 	
RESULT 14		
US-09-252-991A-21105		
; Sequence 21105, Application US/09252991A		
; Patent No. 6551795		
; GENERAL INFORMATION:		
; APPLICANT: Marc J. Rubenfield et al.		
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS		
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS		
; FILE REFERENCE: 107196.136		

352	QY	-----GAGACCACCGCTGAGCCATCGGCCCT-----GCAGCGAAGCAA	3117
		::: :::	
492	Db	AspAlaThrGlnGlnCysAsnLeuThrPheGlyProGluTyrSerValCysProGlyMet	511
		::: :::	
316	QY	GGGACAGCTGCCACACTCAGGGCCAGTAACTACAGTACAGAACGACCACTGAATTCAG	257
		::: :::	
512	Db	AspValCysAlaArgLeuTrpCysAlaValValArgGlnGlnMetValCysLeuThr	531
		::: :::	
256	QY	GCTTCCTCCAAAGCTTTGAAAGTAGACAGTCCAGGCTAT-----	218
		::: :::	
532	Db	LysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgValCysLeuGlnGly	551
		::: :::	
217	QY	AAAACTCTAGAGCACTTCGGTAAGAAGTGTTAAAGTCTTACACA-----	175
		::: :::	
552	Db	LysCysValAspLysThrLysLysLys-TyTyrSerThrSerSerHisGlyAsnTrpG	571
		::: :::	
174	QY	-----AATACATCTTGTAAAACTCAATAAAAATATATATATA	140
		::: :::	
571	Db	ySerTrpGlyProTrpGlyGlnCysSerArgSerCysGlyGlyValcInPheAlaTy	591
		::: :::	
139	QY	TGATATATATAAACTGTGAACATCTTAATACATCGGAACCTGCACACAGGGCGGCCCC	80
		:::	
591	Db	rArgHis-----CysAsn-----AsnProAlaPr	599
		::: :::	
79	QY	TCCCTGGAAACCGTCTCCCTCGCTGGGACACACAGCAATTAGAAGAATTTGTATGAAAT	20
		:::	
599	Db	ArgAsnSerGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSerValTh	619
		:::	
19	QY	ACCA	16
619	Db	rPro	620

Search completed: September 22, 2004, 10:56:32
Job time : 30.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:50:56 ; Search time 54.5 Seconds
(without alignments)

4537.112 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAGCAAGCTGGTA.....GTGCATGGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 2684796

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xmlp

-Q=/cgn21/USFTO.spool_p/CANELLA305B/runat_22092004_112338_14585/app_query.fasta_1.583

-DB=Published Applications_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=200000000 -USER=CANELLA305B @CGN 1.1 225 @runat 22092004_112338_14585

-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOC=100

-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

RESULT 1

US-10-767-701-47126

; Sequence 47126, Application US/10767701

; Publication No. US20040172684A1

; GENERAL INFORMATION: David K.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

; FILE REFERENCE: 38-21(53535)B

; CURRENT APPLICATION NUMBER: US/10767,701

; CURRENT FILING DATE: 2004-01-29

; NUMBER OF SEQ ID NOS: 63128

; SEQ ID NO 47126

; LENGTH: 357

; TYPE: PRT

; ORGANISM: Sorghum bicolor

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(357)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C7832_1.pcp

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
74.5	10.6	357	16	US-10-767-701-47126	Sequence 47126, A																																								
74	10.5	87	14	US-10-029-386-28395	Sequence 28395, A																																								
72.5	10.3	147	15	US-10-264-237-2482	Sequence 2482, Ap																																								
71	10.1	1045	15	US-10-161-927-36	Sequence 36, Appl																																								
71	10.1	2854	15	US-10-085-198-16	Sequence 16, Appl																																								
70.5	10.0	735	16	US-10-437-963-124522	Sequence 124522, A																																								
70.5	10.2	1385	16	US-10-437-963-123966	Sequence 123966, A																																								
70	10.0	1106	16	US-10-437-963-196290	Sequence 196290, A																																								
69.5	9.9	672	16	US-10-437-963-124521	Sequence 124521, A																																								
69.5	10.1	2214	15	US-10-369-493-6737	Sequence 6737, Ap																																								
69	10.0	569	12	US-10-087-192-207	Sequence 207, App																																								
68.5	9.8	135	12	US-10-424-599-158314	Sequence 158314, A																																								
68.5	9.8	391	14	US-10-017-161-2324	Sequence 2324, Ap																																								
68.5	9.8	391	15	US-10-292-798-1970	Sequence 1970, Ap																																								
68	9.6	168	14	US-10-157-031-228	Sequence 228, App																																								
67.5	9.6	116	16	US-10-437-963-198137	Sequence 198137, A																																								
67.5	9.8	118	12	US-10-424-599-203789	Sequence 203789, A																																								
67.5	9.6	141	9	US-09-925-299-787	Sequence 787, App																																								
67.5	9.6	141	10	US-09-925-299-787	Sequence 787, App																																								
67.5	9.8	396	14	US-10-198-070-94	Sequence 94, Appl																																								
67.5	9.6	526	9	US-09-855-154-2	Sequence 2, Appl																																								
67.5	9.8	599	14	US-10-279-733-9	Sequence 9, Appl																																								
67	9.5	525	16	US-10-437-963-177003	Sequence 177003, A																																								
67	9.5	528	14	US-10-253-349-6	Sequence 6, Appl																																								
67	9.7	628	12	US-10-358-283-8	Sequence 8, Appl																																								
67	9.7	752	12	US-10-358-283-6	Sequence 6, Appl																																								
67	9.7	753	12	US-10-358-283-4	Sequence 4, Appl																																								
67	9.7	763	12	US-10-358-283-15	Sequence 15, Appl																																								
67	9.7	930	12	US-10-358-283-2	Sequence 2, Appl																																								
67	9.7	930	12	US-10-358-283-30	Sequence 30, Appl																																								
67	9.7	930	14	US-10-247-685-15	Sequence 15, Appl																																								
67	9.7	1359	16	US-10-408-765A-218	Sequence 218, App																																								
67	9.7	1411	16	US-10-408-765A-237	Sequence 237, App																																								
67	9.5	1638	13	US-10-090-458-2	Sequence 2, Appl																																								
66.5	9.5	124	16	US-10-437-963-191569	Sequence 191569, A																																								
66.5	9.5	147	12	US-10-424-599-165824	Sequence 165824, A																																								
66.5	9.5	259	15	US-10-369-493-20007	Sequence 20007, A																																								
66.5	9.6	450	16	US-10-437-963-149314	Sequence 149314, A																																								
66.5	9.6	491	16	US-10-437-963-157538	Sequence 157538, A																																								
66.5	9.6	1043	9	US-09-946-805-4	Sequence 4, Appl																																								
66.5	9.6	2524	15	US-10-190-115-25	Sequence 25, Appl																																								
66.5	9.6	2524	15	US-10-369-072-25	Sequence 25, Appl																																								
66	9.4	161	14	US-10-156-761-12304	Sequence 12304, A																																								
66	9.4	193	16	US-10-437-963-194777	Sequence 194777, A																																								
66	9.4	294	12	US-10-343-650A-54	Sequence 54, Appl																																								

ALIGNMENTS


```
Publication No. US20030235821A1
GENERAL INFORMATION:
APPLICANT: Zerhusen, Bryan D.
APPLICANT: Kekuda, Ramesh
APPLICANT: Spytek, Kimberly A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Miller, Charles E.
APPLICANT: Hjalt, Tord
APPLICANT: Gerlach, Valerie L.
APPLICANT: Baumgartner, Jason C.
APPLICANT: Guo, Xiaojie
APPLICANT: Gangolli, Bsha A.
APPLICANT: Vernet, Corine
APPLICANT: Padigar, Muralidhara
APPLICANT: Li, Li
APPLICANT: Pena, Carol E.A.
APPLICANT: Gorman, Linda
APPLICANT: Anderson, David W.
APPLICANT: Eisinger, Shlomit R.
APPLICANT: Patturajan, Meera
APPLICANT: Stone, David J.
TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
FILE REFERENCE: 21402-377 D (Cura 677 Other)
CURRENT APPLICATION NUMBER: US/10/161,927
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: 60/295,661
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/295,607
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 60/296,404
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,418
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: 60/296,575
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: 60/297,414
PRIOR FILING DATE: 2001-06-11
PRIOR APPLICATION NUMBER: 60/297,567
PRIOR FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: 60/298,528
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/325,685
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/299,133
PRIOR FILING DATE: 2001-06-18
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 190
SEQ ID NO 36
LENGTH: 1045
TYPE: PRT
ORGANISM: Homo sapiens
US-10-161-927-36

Alignment Scores:
Pred. No.: 46.7 Length: 1045
Score: 71.00 Matches: 24
Percent Similarity: 40.24% Conservative: 9
Best Local Similarity: 29.27% Mismatches: 25
Query Match: 10.11% Indels: 24
DB: 15 Gaps: 5

M64347 (1-385) x US-10-161-927-36 (1-1045)
QY 189 CACTTCTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCT-----TTCAAA 242
Db 225 HsArgLeuHisAnsSerValSerAlaProSerProGlylleProProArgaspPheLys 244
QY 243 GCT-----TGGAGGGAAGCCGTTGAATCA 266
Db 245 SerLeuAlaLeuAlaArgAlaProGlyHisGlyGlyPheTrp---GlnGlyValAlaAla 263
QY 267 GTTGGTTCGTTCTGTACTGTACTGGG-----CCCTGAGTCTGGGCAGCTGTC 314

Publication No. US20040009907A1
GENERAL INFORMATION:
APPLICANT: Alsobrook et al.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-279
CURRENT APPLICATION NUMBER: US/10/085,198
CURRENT FILING DATE: 2002-02-25
PRIOR APPLICATION NUMBER: 60/271,646
PRIOR FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: 60/275,401
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/311,981
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 60/312,858
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/271,840
PRIOR FILING DATE: 2001-02-27
PRIOR APPLICATION NUMBER: 60/277,324
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 60/286,096
PRIOR FILING DATE: 2001-04-21
PRIOR APPLICATION NUMBER: 60/299,695
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/315,614
PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: 60/272,405
PRIOR FILING DATE: 2001-02-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 653
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 2854
TYPE: PRT
ORGANISM: Homo sapiens
US-10-085-198-16

Alignment Scores:
Pred. No.: 55.6 Length: 2854
Score: 71.00 Matches: 25
Percent Similarity: 43.06% Conservative: 6
Best Local Similarity: 34.72% Mismatches: 25
Query Match: 10.11% Indels: 16
DB: 15 Gaps: 3

M64347 (1-385) x US-10-085-198-16 (1-2854)
QY 182 GACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCTGGACTGCTACCTTTCAA 241
Db 819 GluLeuThrLeuLeu-----TrpAspArgThr----- 827
QY 242 AGCTTGGAGGAAGCCGGAATTCAGTTGGTTCGTTCTGTACTGTACTTACTGGGCC----- 295
Db 828 GlyValProGlyGlySerGluLysSerPhePheLeuGluProTyrArgSerSerSer 847
QY 296 ---CTGAGTCTGGGAGCTGTCCTTCCTGCTGCAGGGCCATGGCTCAGGGTGGTCTC 352
Db 848 CysThrSerTyrSerSerCys--LeuGlyCysLeuAlaAspGlnGlyCysGlyTrpCysL 867
QY 353 TCTCTGGGGCCCGCAGTCATGTCATGTCGCCAGAGGT 384

Db 264 GluGlyValGlyCysThrLeuThrGlyAlaTrpArgSerProValProTrpSerGlyThr 283
QY 315 CCTTGCTTGCTGCA-----GGGCCATGGCTCAGGGTGGTCTCTCTCTTGG 359
Db 284 GlyCysValProGlyGlyPheThrValProGlyProArgProAlaProAlaProTrp 303
QY 360 GGCCCA 365
Db 304 GlyPro 305

RESULT 5
US-10-085-198-16
; Sequence 16, Application US/10085198
; Publication No. US20040009907A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-279
; CURRENT APPLICATION NUMBER: US/10/085,198
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/271,646
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/275,401
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,981
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,858
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/271,840
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/277,324
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/286,096
; PRIOR FILING DATE: 2001-04-21
; PRIOR APPLICATION NUMBER: 60/299,695
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/315,614
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/272,405
; PRIOR FILING DATE: 2001-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2854
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-085-198-16
```



```

; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52345200122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2039
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 207
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-207

Alignment Scores:
Pred. No.: 71.5 Length: 569
Score: 69.00 Matches: 40
Percent Similarity: 35.80% Conservative: 18
Best Local Similarity: 24.69% Mismatches: 46
Query Match: 10.00% Indels: 58
DB: 12 Gaps: 10

M64347 (1-385) x US-10-087-192-207 (1-569)
QY 380 CTGGCCACCATGCACTGGGCCCCAGAGACACCACTGAGCATGGCC----- 330
DB 60 LeuAlaArgGlnTyrTrrPheAlaAlaArgAlaHisProGlnPro---ProArgAlaAsp 78
QY 329 -----TGCAGCAAGCAAGGAC 312
DB 79 LeuValSerLeuGlnPheAlaGluLeuPheGlnArgHisPheCysArgGluValArgGlu 98
QY 311 AGCTGCCAGACTCAGGCGCCAGTAACAGTACAGAACCACTGAATTCACGGCTC 252
DB 99 SerLeu-----AlaGlyProProGlyHisAspTyrArgAlaThr-----Ala 112
QY 251 CTTCAAGCTTTGAAAGAGTCCAGCTCCAGGCTATAAACTCTAGAACCATTCGTAAGAA 192
DB 113 ProPro-----ArgProAlaLeuProLysAlaArgSerSer-----Glu 125
QY 191 GTGTGA-----AGTCTACAACAATATACATCTTGTAATAAACTCAATAA 150
DB 126 AspLeuGlyProArgProAlaCysAlaLeuGlnHisLeuArgArgGlyLeuArgGln--- 144
QY 149 ATTATATATAGATATATAATAA-----CTTGTA 120
DB 145 -----LeuPheArgArgSerAlaGlyGluLeuProGlyAlaThrSer 159
QY 119 ACATCTAATACATCGGAACCTGCACAGGCGCGCCCTCCCTCGAAACCGCTCTCCT 60
DB 160 AspThrAsnAspIleAspThrThrAlaAlaSerArgProGlyProAlaArgLysLeuLeu 179
QY 59 GCCTGG-----GACACACAGCAATTAGAAGAAATTTGTATGAATAACACAG 15
DB 180 ProTrpGlyLeuArgGluProProThrGluAlaLeuLysGluValValLeuArgTyrSer 199
QY 14 CTTGCT 9
DB 200 LeuAla 201

RESULT 12
US-10-424-599-158314
; Sequence 158314, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yinhua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 158314
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_113977C.1.pep
US-10-424-599-158314

Alignment Scores:
Pred. No.: 63.5 Length: 135
Score: 68.50 Matches: 24
Percent Similarity: 38.95% Conservative: 6
Best Local Similarity: 31.17% Mismatches: 20
Query Match: 9.76% Indels: 27
DB: 12 Gaps: 4

M64347 (1-385) x US-10-424-599-158314 (1-135)
QY 206 TTCTAGAGTTTATAGCTGCTACTGCTACCTTTCAAGCTTGGAGGGAAGCCGTGAATTC 265
DB 55 PheValSerValValSerPheGlyAlaThr-----Val 65
QY 266 AGTTGGTTCTGCTCTACTGTTTAC-----TGGGCCCTGAGTCTGGGAGCTGT 313
DB 66 SerGluPheValValMetCysValValValPheGlyTrrPileLeuIleTyrGlySerCys 85
QY 314 CCCTTCCTTCCTTCGAGGCCCATGCT-----CAGGGTGGTCTCTCTTTG----- 358
DB 86 ProLeuThrLeuCysAspLeuSerAlaSerGlyArgArgGlyGlyCysPheLeuGluHis 105
QY 359 -----GGGCCAGTCATGTCGCCACAG 382
DB 106 LysValPheGluIleHisMetCysAlaGlyGlySerIleGlyTrrGlnGln 122

RESULT 13
US-10-017-161-2324
; Sequence 2324, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAL, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2324
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (213)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-2324

Alignment Scores:
Pred. No.: 76.5 Length: 391
Score: 68.50 Matches: 29

```

Percent Similarity: 40.00% Conservative: 11
 Best Local Similarity: 29.00% Mismatches: 32
 Query Match: 9.76% Indels: 28
 DB: 14 Gaps: 6

M64347 (1-385) x US-10-017-161-2324 (1-391)

```

QY 73 CCAGGGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATAGATGTTACAGTTTATAT 132
Db 196 ProPheArgSerArgSerCys-----ProLeuLeuSerValProSerAlaLeu 211
QY 133 ATATCTATATATATAATTTATTAGTGTTTTACAGATGATTGTTGTT----- 180
Db 212 Val***Phe-----ValSerCysArgPheValCysCysLeuSerIleAla 226
QY 181 -----AGACTTAACACTTCTTACGCAATGCTTACAGTTTATAGCTTCTCTAC 234
Db 227 SerGlyArgValArgSer-CysValSerCysPhe---SerPheValArgPheValArgLe 245
QY 235 CTTTCAAAGCTTGGAGGAGCGCGTGAATTCAGTTGTTCTGTTCTGTAC----- 283
Db 245 uPhe-----ArgPheIleSerTTPheLeuArgPheArgSerSerVa 259
QY 284 ----TGTTACTGGCCCTGAGTCTGGCAGCTGTCCTTGTGCTGCTGCGAGGCCCATG 337
Db 259 lSerSerPheValArgLeuSerLeuGlyLeuCysSerCysLeuValLeuArgSerVal 278

```

RESULT 14

```

US-10-292-798-1970
; Sequence 1970, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1970
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (213)..(213)
; OTHER INFORMATION: Variable amino acid
US-10-292-798-1970

```

Alignment Scores:
 Pred. No.: 76.5 Length: 391
 Score: 68.50 Matches: 29
 Percent Similarity: 40.00% Conservative: 11
 Best Local Similarity: 29.00% Mismatches: 32
 Query Match: 9.76% Indels: 28
 DB: 15 Gaps: 6

M64347 (1-385) x US-10-292-798-1970 (1-391)

```

QY 73 CCAGGGAGGGCGCGCCCTGTGTCAGGTTCCGATGTTATAGATGTTACAGTTTATAT 132
Db 196 ProPheArgSerArgSerCys-----ProLeuLeuSerValProSerAlaLeu 211
QY 133 ATATCTATATATATAATTTATTAGTGTTTTACAGATGATTGTTGTT----- 180
Db 212 Val***Phe-----ValSerCysArgPheValCysCysLeuSerIleAla 226

```

```

QY 181 -----AGACTTAACACTTCTTACGCAATGCTTACAGTTTATAGCTTCTCTAC 234
Db 227 SerGlyArgValArgSer-CysValSerCysPhe---SerPheValArgPheValArgLe 245
QY 235 CTTTCAAAGCTTGGAGGAGCGCGTGAATTCAGTTGTTCTGTTCTGTAC----- 283
Db 245 uPhe-----ArgPheIleSerTTPheLeuArgPheArgSerSerVa 259
QY 284 ----TGTTACTGGCCCTGAGTCTGGCAGCTGTCCTTGTGCTGCTGCGAGGCCCATG 337
Db 259 lSerSerPheValArgLeuSerLeuGlyLeuCysSerCysLeuValLeuArgSerVal 278

```

RESULT 15

```

US-10-157-031-228
; Sequence 228, Application US/10157031
; Publication No. US2003010890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 228
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (163)..(163)
; OTHER INFORMATION: X = unknown
US-10-157-031-228

```

Alignment Scores:
 Pred. No.: 75.4 Length: 168
 Score: 68.00 Matches: 19
 Percent Similarity: 53.33% Conservative: 5
 Best Local Similarity: 42.22% Mismatches: 12
 Query Match: 9.86% Indels: 9
 DB: 14 Gaps: 2

M64347 (1-385) x US-10-157-031-228 (1-168)

```

QY 356 AGAAGAGACCACCTGAGCCATGCCCTGCAGGCAAGGACAGCTGCCAGACTCA 297
Db 80 ArgArgThrArgProGluPro-GlyProAla---ArgArgGlyAlaArgAlaGluProPr 98
QY 296 GGGCCAGTAAACAGTACAGAACCAACTGAATTCACGGCTTCCCTCCAAGCTTTGAA 237
Db 98 oGlyProProThrLeuGlnAspProProAlaSerHisLeu----- 111
QY 236 AGTAGAGAGTCCA 224
Db 112 -GlySerSerPro 115

```

Search completed: September 22, 2004, 10:58:25
 Job time : 59.5 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:45:25 ; Search time 30 seconds
(without alignments)
2468.916 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAGCAAGCTGTA.....GTGCATGTGCGCAGAGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cpn2.1/USPTO/spool_p/CANEILLA305B/runat_22092004.112336.14518/app.query.fasta_1.583
-DB=PIR_78 -OFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=CANEILLA305B @CGN 1 1 63 @runat_22092004.112336.14518 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	85.5	12.4	674	2 A97490	hypothetical prote
2	74.5	10.6	263	2 T15453	hypothetical prote
3	73.5	10.5	1073	2 I51055	recombination acti
4	72.5	10.3	404	2 T39195	probable amino aci
5	72.5	10.3	404	2 T19922	hypothetical prote
C 6	71	10.3	187	2 S35050	mucin JER28 - huma
7	71	10.1	324	2 C90492	hypothetical prote
8	70.5	10.0	360	2 C81699	conserved hypothet
9	70.5	10.0	1042	2 S42511	RAG-1 protein - ra
10	70.5	10.0	1043	2 B33754	recombination-acti
C 11	70.5	10.2	2257	2 D86483	protein F5J5.19 li
C 12	69.5	10.1	2214	2 T16305	hypothetical prote
13	69	9.8	2090	2 T30075	hypothetical prote
14	69	9.8	2153	2 T30074	hypothetical prote

C 15	68.5	9.9	605	2 G95853	probable pyruvate
16	68	9.7	295	2 S21306	hypothetical prote
17	68	9.7	849	2 T39406	hypothetical prote
18	67.5	9.6	363	2 G71544	hypothetical prote
C 19	67	9.7	326	2 A46676	CD68 homolog macro
20	67	9.5	326	2 H72472	hypothetical prote
C 21	66.5	9.6	1043	2 D84900	hypothetical prote
C 22	66.5	9.6	2524	2 A35844	Xotch protein - Af
C 23	66	9.6	606	2 S43118	finger protein - m
24	65.5	9.3	335	2 T25054	hypothetical prote
C 25	65.5	9.5	863	2 T47038	hypothetical prote
C 26	65.5	9.5	863	2 AD0234	probable fibrillar
C 27	65	9.3	361	2 D84452	hypothetical prote
C 28	64.5	9.3	166	2 D83230	hypothetical prote
29	64.5	9.2	255	2 T41451	very hypothetical
30	64.5	9.2	590	1 S34960	NADH2 dehydrogenas
31	64.5	9.2	603	2 S06059	gene ND1 intron 4
32	64.5	9.2	645	2 T25824	hypothetical prote
C 33	64	9.3	235	2 T31424	C-terminal domain-
C 34	64	9.3	348	2 H70549	probable pdhs prot
C 35	64	9.3	3429	2 T13853	hypothetical prote
36	63.5	9.0	918	2 T58178	glutamate receptor
37	63.5	9.0	949	2 S19808	glutamate receptor
C 38	63.5	9.2	1615	2 JER0372	low density lipopr
39	63	9.0	275	2 T20246	hypothetical prote
40	63	9.0	412	2 F86296	hypothetical prote
C 41	63	9.1	454	2 T35380	probable membrane
C 42	63	9.0	474	2 D91041	probable PTS enzym
43	63	9.0	474	2 G85885	probable PTS enzym
C 44	63	9.1	1970	2 T03284	myoblast city prot
C 45	62.5	9.1	125	2 S20639	Ig heavy chain V r

ALIGNMENTS

RESULT 1

A97490 Hypothetical protein AGR_C_1968 [imported] - Agrobacterium tumefaciens (strain C58, Cerc
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
R:Accession: A97490
A:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: A97490
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-674 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86874.1; PID:GLS156092; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_1968
A:Map position: circular chromosome

Alignment Scores:

Pred. No.: 0.119 Length: 674
Score: 85.50 Matches: 27
Percent Similarity: 43.59% Conservative: 7
Best Local Similarity: 34.62% Mismatches: 25
Query Match: 12.39% Indels: 19
DB: 2 Gaps: 3

M64347 (1-385) x A97490 (1-674)

QY	376	CCACATGCTACCTGGGCCCCAAGAA-----GAG 350
Db	310	ProProAlaLeuGlnProGlnGlnAlaThrProSerGlnProSerAlaAlaThrSerGlu 329
QY	349	ACCCACCTGACCCATGGCCCTGCAGCAAGCAAGGACAGCTGCCACACCTCAG----- 296
Db	330	ThrThrAlaAsnAsnThrProValProSerSerGlyThrAlaAlaProThrAlaAlaGly 349

QY 295 -----GCCCACTAACA-----GTACAGAACGACCAACTGAATTCACGGCT 254
 Db 350 SerThrAlaValAlaProLeuThrSerSerValGlnValThrAlaValGluPheGluGly 369
 QY 253 TCCCTCCAAAGCTTGAAGGTAGCAGTCCAGGCTATAAACTCTAGAAAGCATTTG 200
 Db 370 SerLysIlePheValAlaGlySerAlaProGlyGlySerThrValArgAlaLeu 387

RESULT 2
 T15453
 hypothetical protein CO8A9.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T15453
 R:Latrelle, P.
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of C. elegans cosmid CO8A9.
 A:Reference number: 218353
 A:Accession: T15453
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-263 <LAT>
 A:Cross-references: EMBL:U42844; NID:g1125797; PID:g1125799; PIDN:AAB53817.1; GSPDB:GN00
 A:Experimental source: strain Bristol N2; clone CO8A9
 C:Genetics:
 A:Gene: CESP:CO8A9.3
 A:Map position: X
 A:Introns: 45/3; 79/3; 128/2

Alignment Scores:
 Pred. No.: 2.57 Length: 263
 Score: 74.50 Matches: 29
 Percent Similarity: 39.29% Conservative: 15
 Best Local Similarity: 25.89% Mismatches: 25
 Query Match: 10.61% Indels: 43
 DB: 2 Gaps: 5

M64347 (1-385) x T15453 (1-263)
 QY 28 ACAAATCTCTTAATTCCTGTGTCCAGGAGGAGCGTTCCAGGAGGAGGCGCG 87
 Db 140 SerAsnSerAlaIleIleCysLeu-----GlyArgGlyGln 151

QY 88 CCCTGTGTGCAGGTTCGGATGTTATTAGATGTACAAAGTTTATATATATATATA 147
 Db 152 -----IleGlyIleTyrLeu 156

QY 148 ATTTATTGAGTTTACAGATGT----- 171
 Db 157 ValTyrTyrAlaValGlnLysCysArgPheGluArgGlnSerPheThrLeuPheTyrLys 176

QY 172 ATTTGTTGTAGACTTAACACTTCTTAGCAATGCTTCTAGAGTTTATAGCCTGGACTGC 231
 Db 177 IleCysCysThrLeuIlePheIleValPheMetLeuMetGlu-----CysLeuAsnArg 194

QY 232 TACCTTCAAGCTTGGAGGAGCGGTGAATTCAGTTGCTGCTGCTAGCTGTTACTG 291
 Db 195 TyrLeuAlaAsn-----PheMetLeuThrTyrAsnValLeuLeuThr 208

QY 292 GGCCTGAGTCTGGCAGCTGCCCTGTGCTGCTGCTG 327
 Db 209 ProAlaLysSerArgGlnLeuAsnMetAspCysVal 220

RESULT 3
 I51055
 recombination activating protein - rainbow trout
 C:Species: Oncorhynchus mykiss (rainbow trout)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: I51055
 R:Hansen, J.D.; Kaattari, S.L.
 Immunogenetics 42, 188-195, 1995
 A:Title: The recombination activation gene 1 (RAG1) of rainbow trout (Oncorhynchus mykiss)
 A:Reference number: I51055; MUID:95369845; PMID:7642230

A:Accession: I51055
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1073 <HAN>
 A:Cross-references: EMBL:U15663; NID:9558917; PIDN:AAA80281.1; PID:9558918
 C:Genetics:
 A:Gene: RAG-1
 A:Introns: 489/1
 C:Superfamily: RING finger homology
 C:Keywords: zinc
 F:306-354/Domain: RING finger homology <RNG>

Alignment Scores:
 Pred. No.: 3.17 Length: 1073
 Score: 73.50 Matches: 28
 Percent Similarity: 44.83% Conservative: 11
 Best Local Similarity: 32.18% Mismatches: 28
 Query Match: 10.47% Indels: 20
 DB: 2 Gaps: 4

M64347 (1-385) x I51055 (1-1073)
 QY 59 CAGGAGACGGTTTCCAGGAGGCGCGCTGTGTCAGGTTCCGATGTTATGATG 118
 Db 488 GlnGlyArgGlyPhe-GlyLeuHisProAlaValCysLeu-----AlaIleArgVa 504

QY 119 TTCAAGTTTATATATATATATATATATATATATATATATTTTATGATTTTACAGATGTAT----- 173
 Db 504 LAsnThrPheLeu-----SerCysSerGlnTyrHisLysMetTyrArgTh 519

QY 174 -----TTGTTGTAGACTTAACACTTCTTACGAATGCTTC 208
 Db 519 rValLysAlaThrSerGlyArgGlnIlePheGlnProLeuHisThrLeuArgThrAlaG 539

QY 209 TAGAGTTTATAGCTGACGCTACCTTCAAAAGCTTGGAGGAGGAGCGGTGAATTCAGT 268
 Db 539 uLysGluLeuLeuProGlyTyrHisProPheGlu---TrpGlnProAlaLeuLysSerVa 558

QY 269 TGGTTCGTTCTGTACTGTT 287
 Db 558 LserThrSerCysHisVal 564

RESULT 4
 T39195
 probable amino acid permease - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: T39195
 R:Wedler, H.; Duesterhoeft, A.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, October 1999
 A:Reference number: Z21834
 A:Accession: T39195
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-591 <WED>
 A:Cross-references: EMBL:AL121764; PIDN:CA57428.1; GSPDB:GN00066; SPDB:SPAC9.10
 A:Experimental source: strain 972h-; cosmid c3
 C:Genetics:
 A:Gene: SPDB:SPAC9.10
 A:Map position: 1
 C:Superfamily: choline transport protein

Alignment Scores:
 Pred. No.: 3.74 Length: 591
 Score: 73.00 Matches: 33
 Percent Similarity: 39.34% Conservative: 15
 Best Local Similarity: 27.05% Mismatches: 28
 Query Match: 10.40% Indels: 46
 DB: 2 Gaps: 9

M64347 (1-385) x T39195 (1-591)
 QY 149 TTTATTGAGTTTTTACAAGATGATTTGTTGTAGAC----- 184

Qy	11	CAAGCTGGTATTTCATACAAATCTTCTAATGCTGTG-----TGTCCCAGGCAG	61
		::: ::: ::: ::: ::: :::	
Dd	271	GlnSerGlyMetIleLeuSerPheLeuIleLeuAlaIlePheLeuTrpIleAspTyrHis	290
Qy	62	GGAGACGGTTTCAGGAGGGCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTTA	121
		::: ::: ::: ::: :::	
Dd	291	GlyThrGlyPhe-----Leu	295
Qy	122	CAAGTTTATATATCTATATATAATTTATTGAGTTTTTACAAGATGTATTGTTGTA	181
		::: ::: ::: ::: ::: :::	
Dd	296	IleValTyrLeuPheGlyThrMet---PheIleGluTyrThrTrpAspAla-----	311
Qy	182	GACTTAACACTTCTTACGCAATGCTCTTAGAGTTTTATAGCCTGAGCATGCTACCTTTCAA	241
		::: ::::: ::: ::: ::: ::: :::	
Dd	312	-----ValTyrLeuCysAlaIleGlu-----	318

```

M64347 (1-385) x S35050 (1-187)
Qy 367 ACTGGGCCCCAAGAGAGACACCCTGAGCCCTGAGCGCCTGAGGCACGACGAGGACAGCT 308
    |||||
Db 11 ThrGlyProSerThrAlaThrValThr--GlyProThrGlySerThrAlaThrAla 29
    |||||
Qy 307 GCCCAGACTCAGGCCCCAGTA-----ACAGTACAGACGAAACCAACT 266
    :: |||||
Db 30 SerSerThrGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThrThrProThr 49
    |||||
Qy 265 GAATTCAGGGTCCCTCCCAAGCTTTGAAAGGTAGCAGCTCCAGGCTATAAAACT---CTA 209
    |||||
Db 50 -----ValThrSerSerLysAlaThrProPheSerSerProGlyThrAlaThrAlaLeu 67
    |||||
Qy 208 GAAGCATTCGCT 197
    |||||
Db 68 ProAlaLeuArg 71
    |||||

```

RESULT 7

C90492

Hypothetical protein SS03088 [imported] - Sulfolobus solfataricus

C/Species: Sulfolobus solfataricus

C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C/Accession: C90492

R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.; Jorg, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Jarrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.

Submitted to GenBank, April 2001

Annotation: Sulfolobus solfataricus complete genome.

A:Reference number: A99139
 A:Accession: C90492
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-324 <KUR>
 A:Cross-references: GB:AE006641; NID:g13016505; PIDN:AAK43194.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: SSO3088

Alignment Scores:

Pred. No.: 6.68 Length: 324
 Score: 71.00 Matches: 26
 Percent Similarity: 41.30% Conservativeness: 12
 Best Local Similarity: 28.26% Mismatches: 18
 Query Match: 10.11% Indels: 36
 DB: 2 Gaps: 6

M64347 (1-385) x C90492 (1-324)

Qy 1 GACTTCAAGCAAGCTGGTATTTTCATACAAATCTTCTAATTGCTGTGTGCCAGGCA 60
 Db 253 AspPhelys--AentyrAArgPheHisCysGluLysAspLeuValSerSerAla 271
 Qy 61 GGGAGACGGTTTCCA---GGGAGGGCGCGCCCTGTGTGCAGGTTCCGATGTTATTAGAT 117
 Db 272 -----ProLeuGlyA--gglyTrp----- 277
 Qy 118 GTTACAAGTTTAT 165
 Db 278 -----LeuPheThrGlyLeuLeuLys 284
 Qy 166 -----AGATGATTTTGTGTAGACTTACACTTCTTACCAATGCTTCTAGAGTTTAT 219
 Db 285 SerAsnLysMetValCysAlaAArgileAspThr-GluAsnGluThrLeuLeuAspTyrGln 304
 Qy 220 AGCTGGAGCTGCTACCTTTCACAGCTTGGAGGGAAG 255
 Db 305 AlaLeuGlu---TyrLeuLysLysLeuAlaGlyGlu 315

RESULT 8

C81699 conserved hypothetical protein TC0468 [imported] - Chlamydia muridarum (strain Nigg)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 18-Aug-2000
 A:Accession: C81699
 R:Read, I.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A:Accession: C81699
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-360 <TET>
 A:Cross-references: GB:AE002315; GB:AE002160; NID:g7190506; PIDN:AAF39317.1; PID:g719051
 A:Experimental source: strain Nigg (MoPn)
 C:Genetics:
 A:Gene: TC0468
 C:Superfamily: Chlamydia trachomatis hypothetical protein CT195

Alignment Scores:

Pred. No.: 7.63 Length: 360
 Score: 70.50 Matches: 32
 Percent Similarity: 34.93% Conservativeness: 19
 Best Local Similarity: 21.92% Mismatches: 28
 Query Match: 10.04% Indels: 67
 DB: 2 Gaps: 7

M64347 (1-385) x C81699 (1-360)

Qy 91 TGTGTGAGCTTCCGATGTTATTAGATGTTACAGTTTATAT 132
 Db 178 CysileLeuLeuProfilieThrValValArgValLeuTyrAsnAlaPheArgPheLeu 197

Qy 133 ---ATATCTATATATATAATTTATTGAGTTTTTACAAGATGT----- 171
 Db 198 LeuileProPheTyrIleValPheGlnMetIleArgGlnLeuTyrGlnGluAspLeuPro 217
 Qy 172 -----ATTGTGTG-----AGACTTAACACTTCTTAC----- 198
 Db 218 PheGluGluGlnPheIleCysSerAspIlePheArgGluMetSerArgSerPheValGln 237
 Qy 199 GCAATGCTTCTAGAGTTTATAGCTGAGCTGTACCTTTCAAAGCTTGGAGGGAAGCCG 258
 Db 238 AlaValLysAlaProPheTyrGlyValAlaCysTyrLeuAlaSerLeuTyrGly----- 255
 Qy 259 TGAATTCAGTTGGTTCGTTCTGTACTGTACTGGCCCTGAGTCTGGGACAGCTCTCCCTT 318
 Db 256 -----LeuLeuAsnProLeuSerGlyArgValIleMet 266
 Qy 319 GCT----- 321
 Db 267 AlaserValGluArgAspTrpAsnAsnAspValIleArgSerArgGlyIleTrpGlyIle 286
 Qy 322 -----TGCTGCAGGCCCATGGCTCA----- 342
 Db 287 PheCysGluLysAsnCysLeuLeuGluGlyGlyThrArgSerGlyLeuGlyGlnHis 306
 Qy 343 GGGTGGTCTCTCTTGGG 360
 Db 307 AlaTrpTyrLeuLeuGly 312

RESULT 9

S42511 RAG-1 protein - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
 A:Accession: S42511; S44379
 R:Fuschioti, P.; Harindranath, N.; Mage, R.G.; McCormack, W.T.; Dhanarajan, P.; Roux, K.
 Mol. Immunol. 30, 1021-1032, 1993
 A:Title: Recombination activating genes-1 and -2 of the rabbit: cloning and characterizat
 A:Reference number: S42511; MUID:93354283; PMID:8350872
 A:Accession: S42511
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1042 <FUS>
 A:Cross-references: EMBL:M77666
 R:Fuschioti, P.; Harindranath, N.; Mage, R.G.; McCormack, W.T.; Dhanarajan, P.; Roux, K.
 submitted to the EMBL Data Library, December 1991
 A:Description: Recombination activating genes -1 and -2 of the rabbit: cloning and charac
 A:Reference number: S44379
 A:Accession: S44379
 A:Molecule type: DNA
 A:Residues: 1-607, 'T', 609-1042 <FUS>
 A:Cross-references: EMBL:M77666; NID:g165677; PIDN:AAA03025.1; PID:g165678
 C:Superfamily: RING finger homology
 C:Keywords: DNA binding; zinc
 F:288-336/Domain: RING finger homology <RNG>

Alignment Scores:

Pred. No.: 7.27 Length: 1042
 Score: 70.50 Matches: 23
 Percent Similarity: 45.95% Conservativeness: 11
 Best Local Similarity: 31.08% Mismatches: 21
 Query Match: 10.04% Indels: 19
 DB: 2 Gaps: 4

M64347 (1-385) x S42511 (1-1042)

Qy 84 CCGGCCCTGTGTGCAGGTTCCGATGTTATTAGATGTTACAGTTTATATATATATA 143
 Db 466 ProAlaValCysLeu-----AlaIleArgValAsnThrPheLeu----- 478
 Qy 144 TATAATTTATTGAGTTTTTACAAGATGTAT----- 173
 Db 479 ---SerCysSerGlnTyrHisLysMetTyrArgThrValLysAlaIleThrGlyArgGln 497

A;Residues: 1-2257 <STO>
A;Cross-references: GB:A005172; NID:g12039052; PIDN:AAF18638.2; GSPDB:GN00141
C;Genetics:
A;Gene: F5J5.19
A;Map position: 1
C;Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoy

Alignment Scores:			
Pred. No.:	Score:	Length:	Matches:
Protein:	7.01	2257	30
Percent Similarity:	38.53%		Conservative: 12
Best Local Similarity:	27.52%		Mismatches: 40
Query Match:	10.22%		Indels: 27
DB:	2		Gaps: 4

M64347 (1-385) x D86483 (1-2257)

Qy	364	GGGCCCCCAAGAGACACCCTGTAGC-----CATGCCCTGCAGCG-----AAG	320
Db	1485	GlyProLeuHisGluthrPrilleSerAspGlnTyrLysProLeuGlyThrLeuAspArg	1504
Qy	319	C AAGGGACACTGCCAGACTCCAGGCCCAGTAACAAGTCATGAATTTC	260
Db	1505	GlnArgLeualaalaArgSerAsnThrThrTyrcystyraspPheProLeuAlaPhe	1524
Qy	259	ACGSCITCCCTCCAAGCTTTAAAGGTAGCAGT---CCAGGCTATAAAA--CTTAGAACCA	203
Db	1525	GlyThrAlaLeuGluluLeuLeutPalaserGlnHisproglyVallys-----	1540
Qy	202	T TGGTGTAAGAAGTGTAAAGCTCACAAATAACATCTTTGAAAAA--CAATAAAATTATAT	143
Db	1541	LysProTyr	1543
Qy	142	ATATAGATATATATAAACTTTGTACATCTTAATACATCGGAACCTGCACACAGGCGCGC	83
Db	1544	LysAspThrLeullelueenValVlysluLeuValPheSerLysProGiugLySe-Ser-Gly	1563
Qy	82	CCCTCCCTGGAAACCGTCTCCCTCCCT 56	
Db	1564	ThrSerLeuAspleuValGluArgpro 1572	

RESULT 12

Tl6305
hypothetical protein F40F4.6 - Caenorhabditis elegans
Species: Caenorhabditis elegans
Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
Accession: TL6305
R.Wilson, R.
submitted to the EMBL Data Library, November 1995
Description: The sequence of C. elegans cosmid F40F4.
Reference number: Z18493
Accession: TL6305
Status: preliminary; translated from GB/EMBL/DDBJ
Molecule type: DNA
Residues: 1-2214 <WIL>
Cross-references: EMBL:U40420; NID:g1065513; PIR:g1065514; PIDN:AAA81430.1; C;
Genetics:
Gene: CESP:F40F4.6
Introns: 57/3; 95/1; 302/3; 323/1; 404/1; 468/1; 507/2; 547/3; 595/3;

Alignment Scores:			
Pred. No.:	Score:	Length:	Matches:
Protein:	9.25	2214	21
Percent Similarity:	48.53%		Conservative: 12
Best Local Similarity:	30.88%		Mismatches: 24
Query Match:	10.07%		Indels: 11
DB:	2		Gaps: 4

M64347 (1-385) x TL6305 (1-2214)

Qy	373	CCATCAGCTGGGCCCAAGAGACACCCCTGAGCCATGGCCCTGCAGGCAAGCAA---	317
Db	1969	ProtleThrMetProthrGlnglnThrlealeuthrGlnglyProvalThrGlnglnThr	1981

A;Experimental source: strain 1021, megaplasmid pSymb
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abela, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMB20095
A;Genome: plasmid
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain B

Alignment Scores:
Pred. No.: 12.9 Length: 605
Score: 68.50 Matches: 23
Percent Similarity: 33.02% Conservative: 12
Best Local Similarity: 21.70% Mismatches: 36
Query Match: 9.93% Indels: 35
DB: 2 Gaps: 3

M64347 (1-385) x G95853 (1-605)

QY	325	GGCAAGCAGGACAGCTGCCAGACTCAGGGCCCACTAACAGTACAG-----	278
Db	518	 GlyIleMetAspGlnAlaPheAlaValAspGlyProValIleGlnAlaLeuValAsp	537
QY	277	-----AACGAACCAACTGAATTCACGGCTTCCCTCCAAGCT	242
Db	538	 ArgTyrGluProLeuMetProProLysMetProAlaAspTyrAlaArgAsnPheArgAla	557
QY	241	TTGAAAGGTAGCAGCTCAGGCTATAAACTCTAGAGCATTCGTAAGAGTGTAAAGTC	182
Db	558	 AlaLeuProGluThrProGlyHisGluLysIleGluLysIleLeuArgAsnSerSerVal	577
QY	181	TACAACAATACATCTTGTAAAACTCATAAATTATATATATATATATATAAACTTG	122
Db	578	 GlyArgLys-----	580
QY	121	TAACATCTATAATACATCGGAA-----CCTGCACACAGGCGCGCCCTCCCTGGAA	71
Db	581	 -----ValThrAspGluGluProGlnProHisGluAlaAlaProAspGluAsn	597
QY	70	ACCGTCTCCCTGCCTGGG	53
Db	598	 ThrGlyGluLeuProgly	603

Search completed: September 22, 2004, 10:55:40
Job time : 38 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 22, 2004, 10:37:10 ; Search time 17.5 Seconds

(without alignments)
2291.087 Million cell updates/sec

Title: M64347

Perfect score: 702

Sequence: 1 GACTTCAAGCAAGCTGTA.....GTGCATGTGGCCAGAGGTG 385

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/p/CANELLA305B/runat_22092004_112335_14494/app_query.fasta_1.583
-DB=SwissProt_42 -QFNT=fastan -SUFFIX=runat -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cgi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=CANELLA305B @CEN_1.33 @runat_22092004_112335_14494 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.5	10.5	1073	1 RAG1_ONCMY	Q91187 oncorhynch
2	73	10.4	591	1 YFVA_SCHPO	Q9u18 schizosacch
3	70.5	10.0	1040	1 RAG1_MOUSE	P15919 mus musculu
4	70.5	10.0	1042	1 RAG1_RABIT	P34088 oryctolagus
5	69	10.0	548	1 LNK_MOUSE	O09039 mus musculu
6	67.5	9.8	484	1 IRX5_MOUSE	Q91kq4 mus musculu
7	67	9.7	211	1 VNS3_RSVN	Q01210 rice stripe
8	67	9.7	211	1 VNS3_RSVT	P26658 rice stripe
9	67	9.7	326	1 CD88_MOUSE	P31996 mus musculu
10	67	9.5	528	1 DYR2_HUMAN	Q92630 homo sapien
11	67	9.7	930	1 ATSS_HUMAN	Q9una0 homo sapien
12	67	9.7	1411	1 TCOF_HUMAN	Q13428 homo sapien
13	66.5	9.6	2524	1 NOTC_XENLA	P21783 xenopus lae
14	65.5	9.3	523	1 AAAL_HUMAN	Q9ns82 homo sapien
15	65	9.3	351	1 U204_ARATH	Q929q3 arabidopsis
16	65	9.4	930	1 ATSS_MOUSE	Q9r001 mus musculu
17	64.5	9.2	353	1 VGLM_MCNVX	P52373 murine cyto
18	64.5	9.3	833	1 CASL_MOUSE	O35177 mus musculu

19	64.5	9.2	1233	1 NME3_HUMAN	Q14957 homo sapien
20	64	9.1	228	1 COBS_PYRFU	Q8u400 pyrococcus
21	64	9.3	378	1 LFNG_MOUSE	O09010 mus musculu
22	64	9.3	378	1 LFNG_RAT	Q924t4 rattus norv
23	63.5	9.2	297	1 HXAB_CHICK	P31258 gallus gall
24	63.5	9.0	918	1 GLK1_HUMAN	P39086 homo sapien
25	63.5	9.0	949	1 GLK1_RAT	P22756 rattus norv
26	63	9.1	365	1 LMC1_HUMAN	Q9nzus homo sapien
27	63	9.1	518	1 TEX5_HUMAN	Q99593 homo sapien
28	62.5	9.1	240	1 Y171_HALNI	Q9hs17 halobacteri
29	62.5	9.1	310	1 FMT_GLOVI	Q7njki gloebacteri
30	62.5	8.9	430	1 DGOI_ECOLI	P31457 escherichia
31	62.5	9.1	732	1 POK_DROME	Q01842 drosophila
32	62	8.8	301	1 HMG1_PSEMV	P13703 pseudomonas
33	62	8.8	357	1 YMR2_EBY	P03192 epstein-bar
34	62	8.8	474	1 YFV2_ECOLI	P77272 escherichia
35	62	9.0	1146	1 AGE1_CAEEL	Q94125 caenorhabdi
36	62	8.8	1305	1 CYA9_XENLA	P98999 xenopus lae
37	62	8.8	7073	1 RIAB_CVHSA	P59641 h replicase
38	61.5	8.8	399	1 OYE2_YEAST	Q03558 saccharomyc
39	61.5	8.9	414	1 IP2G_HALNI	Q9hmk9 halobacteri
40	61.5	8.8	530	1 AAAL_MOUSE	Q91mh8 mus musculu
41	61.5	8.8	554	1 YCT3_SCHPO	O59813 schizosacch
42	61.5	8.8	559	1 DACA_HUMAN	Q9Byj9 homo sapien
43	61.5	8.8	875	1 NETR_HUMAN	O56730 homo sapien
44	61.5	8.8	940	1 PTGA_MYCPN	P75569 mycoplasma
45	61.5	8.8	1043	1 RAG1_HUMAN	P15918 homo sapien

ALIGNMENTS

RESULT 1
RAG1_ONCMY
ID RAG1_ONCMY STANDARD; PRT; 1073 AA.
AC Q91187.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V(D)J recombination activating protein 1 (RAG-1).
GN RAG1 OR RAG-1.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Shaeta; TISSUE=testis;
RX MEDLINE=95369845; PubMed=7642230;
RA Hansen J.D., Kaattari S.L.;
RT "The recombination activation gene 1 (RAG1) of rainbow trout
RT (Oncorhynchus mykiss): cloning, expression, and phylogenetic
RT analysis.";
RL Immunogenetics 42:188-195(1995).
CC !- FUNCTION: During lymphocyte development, the genes encoding
CC immunoglobulins and T cell receptors are assembled from variable
CC (V), diversity (D), and joining (J) gene segments. This
CC combinatorial process, known as V(D)J recombination, allows the
CC generation of an enormous range of binding specificities from a
CC limited amount of genetic information. The RAG1/RAG2 complex
CC initiates this process by binding to the conserved recombination
CC signal sequences (RSS) and introducing a double-strand break
CC between the RSS and the adjacent coding segment. These breaks are
CC generated in two steps, nicking of one strand (hydrolysis),
CC followed by hairpin formation (transesterification). RAG1/2 has
CC also been shown to function as a transposase in vitro, and to
CC possess RSS-independent endonuclease activity (end processing) and
CC hairpin opening. RAG1 alone can bind to RSS but stable, efficient
CC binding requires RAG2. All known catalytic activities require the
CC presence of both proteins (By similarity).
CC !- COFACTOR: Binds 1 magnesium or manganese ion per subunit (By
CC similarity).
CC !- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: Thymus, a lower level expression is seen in
 CC the kidney.
 CC -!- DOMAIN: The specific binding to the nonamer RSS motif is mediated
 CC by the nonamer binding domain (NBD) (By similarity).
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
 CC -!- SIMILARITY: Contains 1 nonamer binding (NBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U15663; AA080281.1; -
 DR PIR; I51055; I51055.
 DR HSSP; P15919; LRMD.
 DR InterPro; IPR001841; Znf ring.
 DR Pfam; PF00097; zf-C3HC4_1.
 DR SMART; SM00184; RING_1.
 DR PROSITE; PS00518; ZF_RING_1; 1.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 DR Hydrolase; EC:3.1.1.1; 1.
 KW DNA recombination; Zinc-finger; Metal-binding;
 FT ZN FING 310 349 RING-TYPE.
 FT DN BIND 421 476 NBD.
 FT METAL 533 533 DIVALENT METAL ION (BY SIMILARITY).
 FT METAL 743 743 DIVALENT METAL ION (BY SIMILARITY).
 FT ACT SITE 597 597 BY SIMILARITY.
 SQ SEQUENCE 1073 AA; 121044 MW; C08930808B709B5 CRC64;

Alignment Scores:

Pred. No.:	2,12	Length:	1073
Score:	73.50	Matches:	28
Best local Similarity:	44.83%	Conservative:	11
Query Match:	32.18%	Mismatches:	28
DB:	1	Indels:	20
		Gaps:	4

M64347 (1-385) x RAG1_ONCMY (1-1073)

QY 59 CAGGAGACGGTTTCACGGAGGGCGCGCTGTGCAGGTTCGATGTTATTAGATG 118
 D 488 GlnGlyA:gglyPhe-GlyLeuHisProAlaValCysLeu-----AlaileArgVa 504
 QY 119 TTACAAGTTTATATATATCTATATATATAATTTTGTAGTTTTCACAGATGTAT----- 173
 D 504 lasnThrPheLeu-----SerCysSerClnTyrHisLysMetTyrA:grTh 519
 QY 174 -----TTGTTGTAGACTTAACACTTCTTACGCAATGCTTC 208
 D 519 rVallysAlaThrSerGlyA:gglnPheGlnProLeuHisThrLeuArgThrAlaGl 539
 QY 209 TAGACTTTTATAGCTGCTGCTACCTTCAAAGCTTGGAGGAGCGCGAATTCAGT 268
 D 539 ulysGluLeuLeuprogGlyTyrHisProPheGlu---TrpGlnProAlaLeuLysSerVa 558
 QY 269 TGGTTCGTCTGTACTGT 287
 D 558 lserThrSerCysHisVal 564

RESULT 2

YFYA_SCHPO
 ID YFYA_SCHPO STANDARD; PRT; 591 AA.
 AC QUT18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative amino-acid permease C9.10.
 GN SPAC9.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21848401; PubMed=11859360;
 RX Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 Collins M., Connor R., Cronin A., Davis P., Fiddell T., Fraser A.,
 Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
 Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 Weltjens I., Vanstaels E., Rieger M., Schaefer M., Mueller-Auer S.,
 Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
 Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA The genome sequence of Schizosaccharomycetes pombe.
 RT Nature 415:871-880(2002).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the amino acid permease family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL121764; CAB57428.1; -
 DR PIR; T39195; T39195.
 DR GeneDB; SPombe; SPAC9.10; -
 DR InterPro; IPR002293; AA/rel permease1.
 DR InterPro; IPR004840; AAC permease.
 DR InterPro; IPR004841; Permease region.
 DR Pfam; PF00324; aa_permeases; 1.
 DR PROSITE; PS00218; AMINO-ACID-PERMEASE_1; FALSE_NEG.
 KW Hypothetical protein; Transport; Amino-acid transport;
 KW Transmembrane.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 342 362 POTENTIAL.
 FT TRANSMEM 397 417 POTENTIAL.
 FT TRANSMEM 450 470 POTENTIAL.
 FT TRANSMEM 545 565 POTENTIAL.
 SQ SEQUENCE 591 AA; 65069 MW; FFCD707AC7849D4 CRC64;

Alignment Scores:
 Pred. No.: 2,28 Length: 591
 Score: 73.00 Matches: 33
 Percent Similarity: 29.34% Conservative: 15
 Best local Similarity: 27.05% Mismatches: 28
 Query Match: 10.40% Indels: 46
 DB: 1 Gaps: 9

M64347 (1-385) x YFYA_SCHPO (1-591)
 QY 149 TTTATTGAGTTTTCACAGATCTATTGTTGTAGAC----- 184

FT	MUTAGEN	546	546	D->A: LOSS OF DNA-BINDING.
FT	MUTAGEN	560	560	D->A: LOSS OF DNA-BINDING.
FT	MUTAGEN	597	597	E->Q: IMPAIRED CLEAVAGE.
FT	MUTAGEN	600	600	D->N: LOSS OF CLEAVAGE (BOTH NICKING AND HAIRPIN FORMATION).
FT	MUTAGEN	600	600	D->A: LOSS OF CLEAVAGE AND STRAND TRANSFER ACTIVITIES.
FT	MUTAGEN	708	708	D->N: LOSS OF CLEAVAGE (BOTH NICKING AND HAIRPIN FORMATION).
FT	MUTAGEN	708	708	D->A: LOSS OF CLEAVAGE AND STRAND TRANSFER ACTIVITIES.
FT	MUTAGEN	709	709	E->Q: IMPAIRED CLEAVAGE.
FT	MUTAGEN	709	709	E->A: IMPAIRED CLEAVAGE (DEFECTIVE IN HAIRPIN FORMATION).
FT	MUTAGEN	713	713	E->A,C: IMPAIRED CLEAVAGE (BOTH NICKING AND HAIRPIN FORMATION).
FT	MUTAGEN	719	719	E->Q: IMPAIRED CLEAVAGE.
FT	MUTAGEN	792	792	D->N: IMPAIRED CLEAVAGE.
FT	MUTAGEN	811	811	E->Q: LOSS OF DNA-BINDING.
FT	MUTAGEN	811	811	E->A: IMPAIRED CLEAVAGE.
FT	MUTAGEN	959	959	E->Q: IMPAIRED CLEAVAGE.
FT	MUTAGEN	962	962	E->Q: LOSS OF CLEAVAGE (BOTH NICKING AND HAIRPIN FORMATION).
FT	MUTAGEN	986	986	
FT	HELIX	266	266	
FT	TURN	269	269	
FT	HELIX	273	273	
FT	STRAND	276	276	
FT	HELIX	282	287	
FT	TURN	288	288	
FT	STRAND	289	289	
FT	TURN	291	293	
FT	STRAND	296	296	
FT	STRAND	300	302	
FT	TURN	304	305	
FT	STRAND	308	310	
FT	HELIX	311	320	
FT	TURN	321	321	
FT	STRAND	324	324	
FT	TURN	326	328	
FT	STRAND	331	331	
FT	HELIX	334	336	
FT	STRAND	338	338	
FT	HELIX	342	350	
FT	STRAND	352	354	
FT	TURN	358	359	
FT	STRAND	363	365	
FT	HELIX	366	374	
FT	TURN	375	375	
SQ	SEQUENCE	1040 AA; 119160 MW; 034317A931A0D547 CRC64;		
Alignment Scores:				
Pred. No.:	4.76	Length:	1040	
Score:	70.50	Matches:	23	
Percent Similarity:	45.95%	Conservative:	11	
Best Local Similarity:	31.08%	Mismatches:	21	
Query Match:	10.04%	Indels:	19	
DB:	1	Gaps:	4	
M64347 (1-385) x RAG1_MOUSE (1-1040)				
QY	84	CCGCGCCGTGTCGAGGTCGAGTTATTAGATTACAGTTTATATATATATATATA 143		
DB	464	ProAlaValCysLeu-----AlaIleArgValAsnThrPheLeu----- 476		
QY	144	TATAATTATTGAGTTTTCACAGATGAT----- 173		
DB	477	---SerCysSerGlnTyrHisLysMetTyrArgThrValLysAlaIleThrGlyAsgLIn 495		
QY	174	TTGTGTAGACTTAACACTTCTTACGCAATGCTCTAGAGTTTATAGCCTGGACTGCTA 233		
DB	496	IlePheGlnProLeuHisAlaLeuArgAsnAlaGluLysValLeuLeuProGlyTyrHis 515		

QY	234	CCTTCAAGCTTGAGGAGCCGTCGATTCAGTTGGTTCG 275		
DB	516	ProPheGlu---TrpGlnProLeuLysAsnValSer 528		
RESULT 4				
QY	RAG1_RABIT	STANDARD; PRT; 1042 AA.		
DB	RAG1_RABIT	STANDARD; PRT; 1042 AA.		
QY	01-FEB-1994 (Rel. 28, Created)			
DB	01-FEB-1994 (Rel. 28, Last sequence update)			
QY	28-FEB-2003 (Rel. 41, Last annotation update)			
DB	28-FEB-2003 (Rel. 41, Last annotation update)			
QY	V(D)J recombination activating protein 1 (RAG-1).			
DB	V(D)J recombination activating protein 1 (RAG-1).			
QY	Oryctolagus cuniculus (Rabbit).			
DB	Oryctolagus cuniculus (Rabbit).			
QY	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
DB	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
QY	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
DB	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
QY	NCBI_TaxID=9986;			
DB	NCBI_TaxID=9986;			
QY	SEQUENCE FROM N.A.			
DB	SEQUENCE FROM N.A.			
QY	MEDLINE=93354283; PubMed=8350872;			
DB	MEDLINE=93354283; PubMed=8350872;			
QY	Fuschietti P., Harindranath N., Mage R.G., McCormack W.T.,			
DB	Fuschietti P., Harindranath N., Mage R.G., McCormack W.T.,			
QY	Dhanarajan P., Roux X.H.;			
DB	Dhanarajan P., Roux X.H.;			
QY	"Recombination activating genes-1 and -2 of the rabbit: cloning and			
DB	"Recombination activating genes-1 and -2 of the rabbit: cloning and			
QY	characterization of germline and expressed genes.";			
DB	characterization of germline and expressed genes.";			
QY	Mol. Immunol. 30:1021-1032(1993).			
DB	Mol. Immunol. 30:1021-1032(1993).			
QY	-!- FUNCTION: During lymphocyte development, the genes encoding			
DB	-!- FUNCTION: During lymphocyte development, the genes encoding			
QY	immunoglobulins and T cell receptors are assembled from variable			
DB	immunoglobulins and T cell receptors are assembled from variable			
QY	(V), diversity (D), and joining (J) gene segments. This			
DB	(V), diversity (D), and joining (J) gene segments. This			
QY	combinatorial process, known as V(D)J recombination, allows the			
DB	combinatorial process, known as V(D)J recombination, allows the			
QY	generation of an enormous range of binding specificities from a			
DB	generation of an enormous range of binding specificities from a			
QY	limited amount of genetic information. The RAG1/RAG2 complex			
DB	limited amount of genetic information. The RAG1/RAG2 complex			
QY	initiates this process by binding to the conserved recombination			
DB	initiates this process by binding to the conserved recombination			
QY	signal sequences (RSS) and introducing a double-strand break			
DB	signal sequences (RSS) and introducing a double-strand break			
QY	between the RSS and the adjacent coding segment. These breaks are			
DB	between the RSS and the adjacent coding segment. These breaks are			
QY	generated in two steps, nicking of one strand (hydrolysis),			
DB	generated in two steps, nicking of one strand (hydrolysis),			
QY	followed by hairpin formation (transesterification). RAG1/2 has			
DB	followed by hairpin formation (transesterification). RAG1/2 has			
QY	also been shown to function as a transposase in vitro, and to			
DB	also been shown to function as a transposase in vitro, and to			
QY	possess RSS-independent endonuclease activity (end processing) and			
DB	possess RSS-independent endonuclease activity (end processing) and			
QY	hairpin opening. RAG1 alone can bind to RSS but stable, efficient			
DB	hairpin opening. RAG1 alone can bind to RSS but stable, efficient			
QY	binding requires RAG2. All known catalytic activities require the			
DB	binding requires RAG2. All known catalytic activities require the			
QY	presence of both proteins (By similarity).			
DB	presence of both proteins (By similarity).			
QY	-!- COPACITOR: Binds 1 magnesium or manganese ion per subunit (By			
DB	-!- COPACITOR: Binds 1 magnesium or manganese ion per subunit (By			
QY	similarity).			
DB	similarity).			
QY	-!- SUBCELLULAR LOCATION: Nuclear.			
DB	-!- SUBCELLULAR LOCATION: Nuclear.			
QY	-!- TISSUE SPECIFICITY: Thymus.			
DB	-!- TISSUE SPECIFICITY: Thymus.			
QY	-!- DOMAIN: The specific binding to the nonamer RSS motif is mediated			
DB	-!- DOMAIN: The specific binding to the nonamer RSS motif is mediated			
QY	by the nonamer binding domain (NBD) (By similarity).			
DB	by the nonamer binding domain (NBD) (By similarity).			
QY	-!- SIMILARITY: Contains 1 RING-type zinc finger.			
DB	-!- SIMILARITY: Contains 1 RING-type zinc finger.			
QY	-!- SIMILARITY: Contains 1 nonamer binding (NBD) domain.			
DB	-!- SIMILARITY: Contains 1 nonamer binding (NBD) domain.			
This SWISS-PROT entry is copyright. It is produced through a collaboration				
between the Swiss Institute of Bioinformatics and the EMBL outstation -				
the European Bioinformatics Institute. There are no restrictions on its				
use by non-profit institutions as long as its content is in no way				
modified and this statement is not removed. Usage by and for commercial				
entities requires a license agreement (See http://www.isb-sib.ch/announce/				
or send an email to license@isb-sib.ch).				
QY	EMBL; M77666; AAA03025.1; -			
DB	EMBL; M77666; AAA03025.1; -			
QY	PIR; S42511; S42511.			
DB	PIR; S42511; S42511.			
QY	HSSP; P15919; 1RMD.			
DB	HSSP; P15919; 1RMD.			
QY	InterPro: IPR001841; Znf_ring.			
DB	InterPro: IPR001841; Znf_ring.			
QY	Pfam: PF00097; ZF-C3HC4_1.			
DB	Pfam: PF00097; ZF-C3HC4_1.			
QY	SMART; SM00184; RING_1.			
DB	SMART; SM00184; RING_1.			
QY	PROSITE; PS00518; ZF_RING_1; 1.			
DB	PROSITE; PS00518; ZF_RING_1; 1.			
QY	Hydrolase; Endonuclease; Nuclear protein; DNA-binding;			
DB	Hydrolase; Endonuclease; Nuclear protein; DNA-binding;			
QY	DNA recombination; zinc-finger; Metal-binding.			
DB	DNA recombination; zinc-finger; Metal-binding.			
QY	ZN FING 292 331			
DB	ZN FING 292 331			
QY	FT DNA BIND 391 446			
DB	FT DNA BIND 391 446			
QY	FT METAL 602 602			
DB	FT METAL 602 602			
QY	FT METAL 710 710			
DB	FT METAL 710 710			
QY	ACT_SITE 964 964			
DB	ACT_SITE 964 964			

```

SQ SEQUENCE 1042 AA; 119006 MW; 136C6286C6E22FAE CRC64;

Alignment Scores:
Pred. No.: 4.76 Length: 1042
Score: 70.50 Matches: 23
Percent Similarity: 45.95% Conservative: 11
Best Local Similarity: 31.08% Mismatches: 21
Query Match: 10.04% Indels: 19
DB: 1 Gaps: 4

M64347 (1-385) x RAG1_RABIT (1-1042)
QY 84 CCGGCTGTGTCAGGTCCTGATGTTATTAGATGTTTACAAAGTTTATATATATCTATATA 143
Db 466 ProAlaValCysLeu-----AlaIleArgValAsnThrPheLeu----- 478
QY 144 TATAATTTATGATGTTTTCACAGATGAT- 173
Db 479 ---SerCysSerGlnTyrHisLysMetTyrArgThrValLysAlaIleThrGlyArgGln 497
QY 174 TTGTTGTAGACTTAAACACTTCTTACCAAGTCTTCTAGAGTTTATAGCTGACTGCTA 233
Db 498 IlePheGlnProIleHisAlaLeuA-GASAlaGluLysValLeuLeuProGlyTyrHis 517
QY 234 CTTTCAAAGCTTGGAGGGAAGCGGTAATTCAGTTGGTTTCG 275
Db 518 ProPheGlu---TrpGlnProProLeuLysAsnValSerSer 530

RESULT 5
LNK_MOUSE
ID LNK_MOUSE STANDARD; PRT; 548 AA.
AC O09039;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lymphocyte specific adaptor protein lmk (Signal transduction protein
LNK) (Lymphocyte adaptor protein).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Joquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length

```

```

RT human and mouse cDNA sequences.":
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
CC -!- FUNCTION: Links T-cell receptor activation signal to phospholipase
CC C-gamma-1, GRB-2 and phosphatidylinositol 3-kinase (by
CC similarity).
CC -!- PTM: Tyrosine phosphorylated (By similarity).
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U89992; AAB58580.2; -.
CC EMBL; U89993; AAB58581.1; -.
CC EMBL; BC006759; AAO6759.1; -.
CC HSSP; P29353; 1MIL.
CC MGD; MGI:993598; Lnk.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000980; SH2.
CC Pfam; PF00017; SH2; 1.
CC PRINTS; PR00401; SH2DOMAIN.
CC ProDom; PD000093; SH2; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00252; SH2; 1.
CC PROSITE; PS50001; SH2; 1.
CC Phosphorylation; SH2 domain. PH.
FT DOMAIN 168 279
FT DOMAIN 336 434
SQ SEQUENCE 548 AA; 60487 MW; D21DCB46185962B8 CRC64;

Alignment Scores:
Pred. No.: 6.7 Length: 548
Score: 69.00 Matches: 40
Percent Similarity: 35.80% Conservative: 18
Best Local Similarity: 24.69% Mismatches: 46
Query Match: 10.00% Indels: 58
DB: 1 Gaps: 10

M64347 (1-385) x LNK_MOUSE (1-548)
QY 380 CTGGCCCATGCTAGTGGCCCAAGAGACACCTGAGCCATGGCCC----- 330
Db 39 LeuAlaArgGlnTyrTrpLeuPheAlaArgAlaHisProGlnPro---ProArgAlaAsp 57
QY 329 -----TGACGCAACCAAGGAC 312
Db 58 LeuValSerLeuGlnPheAlaGluLeuPheGlnArgHisPheCysArgGluValArgGlu 77
QY 311 AGCTGCCAGACTCAGGCCCCAGTAAACAGTACAGAACCAACCACTGAATTCACGGCTTC 252
Db 78 SerLeu-----AlaGlyProGlyHisAspTyrArgAlaThr-----Ala 91
QY 251 CTTCAAGCTTTGAAAGTAGTACGCTCAGGCTATATAAACTCTAGAGCATTCGGTAAGAA 192
Db 92 ProPro-----ArgProAlaLeuProLysAlaArgSerSer-----Glu 104
QY 191 GTGTTA-----AGTCTACCAAAATACATCTTGTAAAAAACTCAATAA 150
Db 105 AspLeuGlyProArgProAlaCysAlaLeuGlnHisLeuArgArgGlyLeuArgGln--- 123
QY 149 ATTATATATATAGATATATATAA-----CTTGTA 120
Db 124 -----LeuPheArgArgSerAlaGlyGluLeuProGlyAlaThrSer 138
QY 119 ACATCTTAATACATCGGACCCCTGCACACAGGCGCCCTCCCTCGGAACCGTCTCCCT 60
Db 139 AspThrAsnAspIleAspThrThrAlaAlaSerArgProGlyProAlaArgLysLeuLeu 158

```

QY 59 GCCTGG-----GACACAGCAATTAAGAGAAATTGTATGAATACGAC 15
 Db 159 ProTPGlyLeuArgGluProThrGluAlaLeuGlyGluValValLeuArgTySer 178
 QY 14 CTGTGT 9
 Db 179 LeuAla 180

RESULT 6
 IRX5_MOUSE
 ID IRX5_MOUSE STANDARD; PRT; 484 AA.
 AC Q9JKO4; Q9JLL5;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Iroquois-class homeodomain protein IRX-5 (Iroquois homeobox protein
 DE 5).
 GN IRX5.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20284900; PubMed=10822268;
 RA Bosse A., Stoykova A., Nieselt-Struwe K., Chowdhury K., Copeland N.G.,
 RA Jenkins N.A., Gruss P.;
 RT "Identification of a novel mouse Iroquois homeobox gene, Irx5, and
 RT chromosomal localisation of all members of the mouse Iroquois gene
 RT family";
 RL Dev. Dyn. 218:160-174 (2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., Fujita M., Gariboldi M.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Yashaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 403:685-690 (2001).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22380257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smalilus D.E.,
 RA Schenck A., Schein J.E., Jones S.J.M., Marra W.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]

RP SEQUENCE OF 1-442 FROM N.A.
 RX MEDLINE=20171054; PubMed=10704856;
 RA Cohen D.R., Cheng C.W., Cheng S.H., Hui C.C.;
 RT "Expression of two novel mouse Iroquois-class homeobox genes during
 RT neurogenesis";
 RL Mech. Dev. 91:317-321 (2000).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: Belongs to the TALE/IRO homeobox family.
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
 CC frameshift in position 430.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF230074; AAF42871.1; -;
 DR EMBL; AK004747; BAB23528.1; -;
 DR EMBL; BC051959; AAH51959.2; -;
 DR EMBL; BC056994; AAH56994.1; -;
 DR EMBL; AF165985; AAF63955.1; ALT_FRAME.
 DR HSSP; P41778; 1DU6.
 DR TRANSFAC; T04283; -;
 DR MGD; MGI:1859086; Irx5.
 DR InterPro; IPR001356; Homeobox.
 DR InterPro; IPR003893; Iroquois_homeo.
 DR Pfam; PF00046; Homeobox; 1.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR SMART; SM00548; IRO; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR Homeobox; DNA-binding; Nuclear protein.
 KW DNA_BIND 112 174 HOMEBOX (TALE-TYPE).
 FT DOMAIN 72 78 POLY-ALA.
 FT DOMAIN 185 191 POLY-GLU.
 FT DOMAIN 319 327 POLY-PRO.
 FT DOMAIN 351 354 POLY-GLY.
 SQ SEQUENCE 484 AA; 50755 MW; 4C9AB3ED9992EDC9 CRC64;

Alignment Scores:
 Pred. No.: 9.93 Length: 484
 Score: 67.50 Matches: 31
 Percent Similarity: 31.93% Conservative: 7
 Best Local Similarity: 26.05% Mismatches: 36
 Query Match: 9.78% Indels: 45
 DB: 1 Gaps: 6

M64347 (1-385) x IRX5_MOUSE (1-484)
 QY 376 CCACCATGCTACTGGGCCCCAAGAGAGACACACCTTG-----AGCCATGGC 332
 Db 361 ProProCysProGlyProMetGlyGlyGlnThrLeuGlySerArgAlaSerProAla 380
 QY 331 CTGTGAGGCAAGCAGGACAGCTGCCAGACTAGGGCCCA---GTAACAGTACAGAAC 275
 Db 381 ProAlaProAlaArgSerProSerAlaGlnCysProPheProGlyGlyThrValLeuSer 400
 QY 274 GAACCAACTGATTCACCGCTTCCCTCCCAAGCTTTGAAAGGTAGCAGTCCAGGCTATAAA 215
 Db 401 ArgProLeuTyTyThrAlaProPheTy-----ProGlyTyThr 414

RP SEQUENCE FROM N.A.
RX MEDLINE=98148094; PubMed=9479000;
RA Li A.C., Guidez F.R.B., Collier J.G., Glass C.K.;
RT "The macroalgalin promoter directs high levels of transcriptional
activity in macrophages dependent on combinatorial interactions
between PU.1 and c-Jun.";
RL J. Biol. Chem. 273:5389-5399(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=98317533; PubMed=9653646;
RA Jiang Z., Shih D.M., Xia Y.R., Luis A.J., de Beer F.C.,
RA de Villiers W.J.S., van der Westhuizen D.R., de Beer M.C.;
RT "Structure, organization, and chromosomal mapping of the gene encoding
macrostalinal, a macrophage-restricted protein.";
RL Genomics 50:199-205(1998).
CC -!- FUNCTION: Could play a role in phagocytic activities of tissue
macrophages, both in intracellular lysosomal metabolism and
extracellular cell-cell and cell-pathogen interactions. Bind to
tissue- and organ-specific lectins or selectins, allowing homing
of macrophage subsets to particular sites. Rapid recirculation of
CD68 from endosomes, lysosomes to the plasma membrane may allow
macrophages to crawl over selectin bearing substrates or other
cells.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOSOMAL OR
LYSOSOMAL (LONG VARIANT) AND TO A LESSER EXTENT ON THE CELL
SURFACE (SHORT VARIANT).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P31996-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P31996-2; Sequence=VSP_003043;
CC -!- TISSUE SPECIFICITY: Expressed in tissue macrophages and to a
lesser extent in dendritic cells.
CC -!- PTM: N- and O-glycosylated.
CC -!- SIMILARITY: Belongs to the LAMP family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X68273; CAA48334.1; -;
DR EMBL; AF039399; AAC40056.1; -;
DR EMBL; AF022651; AAC40151.1; -;
DR F01; A46676; A46676.
DR MGI; MGI:88342; Cd68.
DR Pfam; PF01299; LAMP; 1.
DR InterPro; IPR002000; LAMP.
DR PRINTS; PR00336; LYSASOCTMP.
DR PROSITE; PS00311; LAMP 2; 1.
KW Transmembrane; Glycoprotein; Signal; Lysosome; Repeat; Antigen;
KW Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 326
FT DOMAIN 21 291
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 292 316
FT DOMAIN 317 326
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 21 109
FT MUCIN-LIKE.
FT DOMAIN 110 121
FT PRO-RICH (HINGE).
FT REPEAT 44 49
FT REPEAT 50 64
FT REPEAT 65 72
FT REPEAT 73 88
FT REPEAT 139 177
FT DISULFID 139 177
FT DISULFID 249 286
FT DISULFID 60 60
FT CARBOHYD 84 84
FT CARBOHYD 97 97

FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 319 326 Missing (in isoform Short).
SQ SEQUENCE 326 AA; 34818 MW; AB7203A9A7EA47BA CRC64;
/FTID=VSP_003043.
Alignment Scores:
Pred. No.: 10.9 Length: 326
Score: 67.00 Matches: 32
Percent Similarity: 35.85% Conservative: 6
Best Local Similarity: 30.19% Mismatches: 33
Query Match: 9.71% Indels: 35
DB: 1 Gaps: 6
M64347 (1-385) x CD68_MOUSE (1-326)
QY 367 ACTGGGCCCCAAGAGAGACACCCCTGAGCCATGGCCCTGCGAGGCAAGCAAGGACGCT 308
Db ThrThrSerHisArgProThrThrSerHisGlyAsnVal-----ThrVal 63
QY 307 GCCCAGACTCAGGGCCCGAGTACAGTACAGAACCAACCACTGAATTCACGGCTTCCCTC 248
Db HisThrSerSerGlyProThrThrValThrHisAsnProAlaThrThr----- 80
QY 247 CAAGCTTTGAAAGGTAGCAGTCCAGGCTAATAAACTCTAGAAGCATTCGGTAAAGTGT 188
Db 81 -----SerHisGlyAsnAlaThrIle-SerHisAla-----ThrVa 92
QY 187 TAAGTCTACAACAAT-----ACATCTGTAAAAAATCAATAAATATATATATAG 137
Db 92 lserProThrThrAsnGlyThrAlaThrSerProArgSerThr----- 107
QY 136 ATATATATAAACTTGTAACATCTAATAACATCGGAACCTGCACAGCGCGGCCCTCC 77
Db 108 -----ValGlyProHisProGlyProProPro 117
QY 76 CTGGAACCGCTCTCCC 61
Db 117 O---SerProSerPro 121
RESULT 10
DVR2 HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.-).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eirnbter K., Tejedor F.J.,
RA Joost H.-G.;
RT "Sequence characteristics, subcellular localization, and substrate
specificity of DYRK-related kinases, a novel family of dual
specificity protein kinases.";
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

M64347 (1-385) x DYR2_HUMAN (1-528)

QY 117 TGTTCACAGATTATATATATCTATATATATAATTTATTGAGTTTTCACAGATGTTATTTG 176

DB 300 CystTyGluHisGlnArgValTyThrTyTyrIleGlnSerArgPheTyArgAlaProG 319

QY 177 TTGTAGACTTAACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCT 236

DB 319 uValleLeuGlyAlaArgTyGlyMetProIleAspMetTrpSerLeuGlyCysIleLe 339

QY 237 TTCAAAGCTTGGAGGAAGCCGCTGAATTCAGTTTGGTTTCGTTCTGTACTGTTTACTGGGCC 296

DB 339 uAlaGluLeu-----LeuThrGlyTyProLeuLeuProGlyG 352

QY 297 TGAGTCTGGGAGCTGCCCTTGTTCGCTG 327

DB 352 uAspGluGlyAsp---GlnLeuAlaCysMet 361

RESULT 11

ATSS5 HUMAN

ID_ATSS5 HUMAN STANDARD; PRT; 930 AA.

AC Q9URAO; Q9UKP2; 930

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2) (ADMP-2) (ADAM-TS 11).

DE DE (ADMP-2) (ADAM-TS 11).

GN ADAMTS5 OR ADMP2 OR ADAMTS11.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_Taxid=9606;

RP [1]

RN SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=99367476; PubMed=10438522;

RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H., Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M., Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K., Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H., Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F., Arner E.C., Burn T.C.,

RA "Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family.";

RT J. Biol. Chem. 274:23443-23450(1999).

RN [2]

RN SEQUENCE FROM N.A.

RP MEDLINE=20289799; PubMed=10830953;

RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Teraki Y., Choi D.-K., Groner Y., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar U., Schillhabel M., Schudy A., Zimmermann W., Reichwald K., Rump A., Kudo J., Shibuya K., Kawasaki K., Asakawa S., Rosenthal A., Kudo J., Nagamine K., Mitsuyama S., Antonakis S.E., Shintani A., Sasaki T., Nagamine K., Hornischer K., Brandt P., Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Antonakis S.E., Scarfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehkrach H., Reinhardt R., Yaspo M.-L.;

RA "The DNA sequence of human chromosome 21.";

RT Nature 405:311-319(2000).

RN [3]

RN SEQUENCE OF 413-930 FROM N.A.

RP TISSUE=Fetal brain;

RC MEDLINE=9935124; PubMed=10464288;

RA Hurskainen T.U., Hirchata S., Seidlin M.P., Apte S.S.;

RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases.";

RT J. Biol. Chem. 274:25555-25563(1999).

CC -1- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may be


```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96154183; PubMed=8563749;
RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RT "Positional cloning of a gene involved in the pathogenesis of
RT Treacher Collins syndrome."
RL Nat. Genet. 12:130-136(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97228900; PubMed=9074926;
RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,
RA Dixon M.J.;
RT "Identification of the complete coding sequence and genomic
RT organization of the Treacher Collins syndrome gene."
RL Genome Res. 7:223-234(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250498; PubMed=9096354;
RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA Ashley J.A., Lovett M., Jabs E.W.;
RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
RT mutations in Treacher Collins syndrome throughout its coding
RT region."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN [4]
RP VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
RP ARG-53.
RX MEDLINE=97195537; PubMed=9042910;
RA Edwards S.J., Gladwin A.J., Dixon M.J.;
RT "The mutational spectrum in Treacher Collins syndrome reveals a
RT predominance of mutations that create a premature-termination
RT codon."
RL Am. J. Hum. Genet. 60:515-524(1997).
CC -1- DISEASE: Defects in TCOF1 are the cause of Treacher Collins
CC syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
CC of craniofacial development that occurs with an incidence of
CC 1/50,000 live births. The clinical features of TCS are bilaterally
CC symmetrical and include: (1) abnormalities of the external ears,
CC atresia of the external ear canals, and malformation of the middle
CC ear ossicles, which may result in conductive hearing loss; (2)
CC lateral downward sloping of palpebral fissures, frequently with
CC colobomas of the lower eyelids; (3) hypoplasia of the mandible and
CC zygomatic complex; (4) cleft palate.
CC -1- SIMILARITY: Contains 1 Lish domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40847; AAC50903.1; -
DR EMBL; U76366; AAC51181.1; -
DR EMBL; U84664; AAC51185.1; -
DR EMBL; U84640; AAC51185.1; JOINED.
DR EMBL; U84641; AAC51185.1; JOINED.
DR EMBL; U84642; AAC51185.1; JOINED.
DR EMBL; U84643; AAC51185.1; JOINED.
DR EMBL; U84644; AAC51185.1; JOINED.
DR EMBL; U84645; AAC51185.1; JOINED.
DR EMBL; U84646; AAC51185.1; JOINED.
DR EMBL; U84647; AAC51185.1; JOINED.
DR EMBL; U84648; AAC51185.1; JOINED.
DR EMBL; U84649; AAC51185.1; JOINED.
DR EMBL; U84650; AAC51185.1; JOINED.
DR EMBL; U84651; AAC51185.1; JOINED.
DR EMBL; U84652; AAC51185.1; JOINED.
DR EMBL; U84653; AAC51185.1; JOINED.
DR EMBL; U84654; AAC51185.1; JOINED.

```

```

DR EMBL; U84555; AAC51185.1; JOINED.
DR EMBL; U84556; AAC51185.1; JOINED.
DR EMBL; U84557; AAC51185.1; JOINED.
DR EMBL; U84558; AAC51185.1; JOINED.
DR EMBL; U84559; AAC51185.1; JOINED.
DR EMBL; U84560; AAC51185.1; JOINED.
DR EMBL; U84561; AAC51185.1; JOINED.
DR EMBL; U84562; AAC51185.1; JOINED.
DR EMBL; U84563; AAC51185.1; JOINED.
DR EMBL; U79659; AAB40722.1; -
DR EMBL; U79645; AAB40722.1; JOINED.
DR EMBL; U79646; AAB40722.1; JOINED.
DR EMBL; U79647; AAB40722.1; JOINED.
DR EMBL; U79648; AAB40722.1; JOINED.
DR EMBL; U79649; AAB40722.1; JOINED.
DR EMBL; U79650; AAB40722.1; JOINED.
DR EMBL; U79651; AAB40722.1; JOINED.
DR EMBL; U79652; AAB40722.1; JOINED.
DR EMBL; U79653; AAB40722.1; JOINED.
DR EMBL; U79654; AAB40722.1; JOINED.
DR EMBL; U79655; AAB40722.1; JOINED.
DR EMBL; U79656; AAB40722.1; JOINED.
DR EMBL; U79657; AAB40722.1; JOINED.
DR EMBL; U79658; AAB40722.1; JOINED.
DR Genew; HGNC:11654; TCOF1.
DR MIM; 606847; -
DR MIM; 154500; -
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0005215; P:transporter activity; TAS.
DR GO; GO:0001501; P:skeletal development; TAS.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR003993; treacle.
DR Pfam; PF03546; treacle; 3.
DR PRINTS; PR01503; TREACLE.
DR SMART; SM00667; Lish; 1.
DR PROSITE; PS50896; Lish; 1.
DR Disease mutation; Polymorphism.
KW DOMAIN 6 38
FT DOMAIN 89 97 POLY-GLU.
FT DOMAIN 204 207 POLY-SER.
FT DOMAIN 616 619 POLY-SER.
FT DOMAIN 919 924 POLY-SER.
FT DOMAIN 1285 1289 POLY-LYS.
FT DOMAIN 1375 1386 POLY-LYS.
FT DOMAIN 1398 1405 POLY-LYS.
FT VARIANT 53 53 W -> R (in TCS).
FT VARIANT 439 439 P -> L.
FT VARIANT 810 810 /FTID=VAR_005631.
FT VARIANT 1313 1313 A -> V.
FT VARIANT 1313 1313 /FTID=VAR_005632.
FT VARIANT 1355 1355 A -> V (in dbSNP:15251).
FT VARIANT 1355 1355 /FTID=VAR_005633.
FT VARIANT 1355 1355 D -> G.
FT VARIANT 1312 1312 /FTID=VAR_005634.
FT CONFLICT 1312 1312 K -> Q (IN REF 2).
SQ SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

```

```

Alignment Scores:
Pred. No.: 12.7 Length: 1411
Score: 67.00 Matches: 23
Percent Similarity: 43.08% Conservative: 5
Best Local Similarity: 35.38% Mismatches: 27
Query Match: 9.71% Indels: 10
DB: 1 Gaps: 4

```

M64347 (1-385) x TCOF_HUMAN (1-1411)

Qy 376 CCACCATGCACTGGGCCCAAGACAGACACCCCTGAGCCATGGCCCTGCA----- 326

Db 752 ProProValArgAsnProGlnAsnSerThrValLeuAlaArgGlyProAlaSerValPro 771

Qy 325 -----GGCAAGCAA---GGGACAGCTGCCAGACTCAG---GGCCCCAGTAACA----- 284

Dd 772 SerValGlyLeuAlaValAlaThrAlaGlnThrGlyProGluGluAspSer 791
 Qy 283 ---GTACAGACGACCACTGATTCACGGCTCCCTCCAGCTTGAAGGTAGCAGT 227
 Dd 792 GlySerSerGluGluGluSerAspSerGluGluGluGluThrLeuAlaGlnAlaLys 811
 Qy 226 CCAGGCTATAAACT 212
 Dd 812 ProSerGlyLysThr 816
 RESULT 13
 ID NOTC_XENLA STANDARD; PRT; 2524 AA.
 AC P21783;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch protein homolog precursor (NOTCH protein).
 GN NOTCH.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_taxid=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90385285; PubMed=2402639;
 RA Coffman C., Harris W., Kintner C.;
 RT "Xor-ch, the Xenopus homolog of Drosophila notch.";
 RL Science 249:1438-1441(1990).
 RN [2]
 RP REVISIONS TO 1759-1782.
 RA Kintner C.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DEVELOPMENTAL STAGE: Expressed almost uniformly in early embryos.
 CC -!- SIMILARITY: Belongs to the NOTCH family.
 CC -!- SIMILARITY: Contains 36 EGF-like domains.
 CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
 CC -!- SIMILARITY: Contains 6 ANK repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M33874; AA02039.1; --
 DR HSPF; P00740; IEDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl_s.
 DR InterPro; IPR000742; EGF 2.
 DR InterPro; IPR001881; EGF Ca.
 DR InterPro; IPR001438; EGF II.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR008297; Notch.
 DR InterPro; IPR000800; Notch_dom.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PIRSF; PIRSF002279; Notch; 1.
 DR PRINTS; PR00010; EGFELOOD.
 DR PRINTS; PR00011; EGFELAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF CA; 24.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS02097; ANK_REPEAT; 1.
 DR PROSITE; PS50086; ANK_REPEAT; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS00026; EGF_3; 36.
 DR PROSITE; PS01187; EGF_CA; 21.
 DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
 FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1729 1750 POTENTIAL.
 FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 57 EGF-LIKE 1.
 FT DOMAIN 58 99 EGF-LIKE 2.
 FT DOMAIN 102 140 EGF-LIKE 3.
 FT DOMAIN 141 177 EGF-LIKE 4.
 FT DOMAIN 179 215 EGF-LIKE 5.
 FT DOMAIN 217 254 EGF-LIKE 6.
 FT DOMAIN 256 292 EGF-LIKE 7.
 FT DOMAIN 294 332 EGF-LIKE 8.
 FT DOMAIN 334 370 EGF-LIKE 9.
 FT DOMAIN 371 409 EGF-LIKE 10.
 FT DOMAIN 411 449 EGF-LIKE 11.
 FT DOMAIN 451 487 EGF-LIKE 12.
 FT DOMAIN 489 525 EGF-LIKE 13.
 FT DOMAIN 527 563 EGF-LIKE 14.
 FT DOMAIN 565 600 EGF-LIKE 15.
 FT DOMAIN 602 638 EGF-LIKE 16.
 FT DOMAIN 640 675 EGF-LIKE 17.
 FT DOMAIN 677 713 EGF-LIKE 18.
 FT DOMAIN 715 750 EGF-LIKE 19.
 FT DOMAIN 752 788 EGF-LIKE 20.
 FT DOMAIN 790 826 EGF-LIKE 21.
 FT DOMAIN 828 866 EGF-LIKE 22.
 FT DOMAIN 868 904 EGF-LIKE 23.
 FT DOMAIN 906 942 EGF-LIKE 24.
 FT DOMAIN 944 980 EGF-LIKE 25.
 FT DOMAIN 982 1018 EGF-LIKE 26.
 FT DOMAIN 1020 1056 EGF-LIKE 27.
 FT DOMAIN 1058 1094 EGF-LIKE 28.
 FT DOMAIN 1096 1142 EGF-LIKE 29.
 FT DOMAIN 1144 1180 EGF-LIKE 30.
 FT DOMAIN 1182 1218 EGF-LIKE 31.
 FT DOMAIN 1220 1264 EGF-LIKE 32.
 FT DOMAIN 1266 1304 EGF-LIKE 33.
 FT DOMAIN 1306 1346 EGF-LIKE 34.
 FT DOMAIN 1347 1383 EGF-LIKE 35.
 FT DOMAIN 1386 1424 EGF-LIKE 36.
 FT REPEAT 1441 1478 LIN/NOTCH 1.
 FT REPEAT 1479 1520 LIN/NOTCH 2.
 FT REPEAT 1521 1560 LIN/NOTCH 3.
 FT REPEAT 1876 1919 ANK 1.
 FT REPEAT 1924 1953 ANK 2.
 FT REPEAT 1957 1987 ANK 3.
 FT REPEAT 1991 2020 ANK 4.
 FT REPEAT 2024 2053 ANK 5.
 FT REPEAT 2057 2086 ANK 6.
 FT DISULFID 22 35 BY SIMILARITY.
 FT DISULFID 29 45 BY SIMILARITY.
 FT DISULFID 47 56 BY SIMILARITY.
 FT DISULFID 62 74 BY SIMILARITY.
 FT DISULFID 68 87 BY SIMILARITY.
 FT DISULFID 89 98 BY SIMILARITY.
 FT DISULFID 106 117 BY SIMILARITY.
 FT DISULFID 111 128 BY SIMILARITY.
 FT DISULFID 130 139 BY SIMILARITY.
 FT DISULFID 145 156 BY SIMILARITY.
 FT DISULFID 150 165 BY SIMILARITY.
 FT DISULFID 167 176 BY SIMILARITY.
 FT DISULFID 183 194 BY SIMILARITY.
 FT DISULFID 188 203 BY SIMILARITY.
 FT DISULFID 205 214 BY SIMILARITY.
 FT DISULFID 221 232 BY SIMILARITY.

```

FT D:SULFID 226 BY SIMILARITY. 242
FT D:SULFID 244 BY SIMILARITY. 253
FT D:SULFID 260 BY SIMILARITY. 271
FT D:SULFID 265 BY SIMILARITY. 280
FT D:SULFID 282 BY SIMILARITY. 291
FT D:SULFID 298 BY SIMILARITY. 311
FT D:SULFID 305 BY SIMILARITY. 320
FT D:SULFID 322 BY SIMILARITY. 331
FT D:SULFID 338 BY SIMILARITY. 349
FT D:SULFID 343 BY SIMILARITY. 358
FT D:SULFID 360 BY SIMILARITY. 369
FT D:SULFID 375 BY SIMILARITY. 386
FT D:SULFID 380 BY SIMILARITY. 397
FT D:SULFID 399 BY SIMILARITY. 408
FT D:SULFID 415 BY SIMILARITY. 428
FT D:SULFID 422 BY SIMILARITY. 437
FT D:SULFID 439 BY SIMILARITY. 448
FT D:SULFID 455 BY SIMILARITY. 466
FT D:SULFID 460 BY SIMILARITY. 475
FT D:SULFID 477 BY SIMILARITY. 486
FT D:SULFID 493 BY SIMILARITY. 504
FT D:SULFID 498 BY SIMILARITY. 513
FT D:SULFID 515 BY SIMILARITY. 524
FT D:SULFID 531 BY SIMILARITY. 542
FT D:SULFID 536 BY SIMILARITY. 551
FT D:SULFID 553 BY SIMILARITY. 562
FT D:SULFID 569 BY SIMILARITY. 579
FT D:SULFID 574 BY SIMILARITY. 588
FT D:SULFID 590 BY SIMILARITY. 599
FT D:SULFID 606 BY SIMILARITY. 617
FT D:SULFID 611 BY SIMILARITY. 626
FT D:SULFID 628 BY SIMILARITY. 637
FT D:SULFID 644 BY SIMILARITY. 654
FT D:SULFID 649 BY SIMILARITY. 663
FT D:SULFID 665 BY SIMILARITY. 674
FT D:SULFID 681 BY SIMILARITY. 692
FT D:SULFID 686 BY SIMILARITY. 701
FT D:SULFID 703 BY SIMILARITY. 712
FT D:SULFID 719 BY SIMILARITY. 729
FT D:SULFID 724 BY SIMILARITY. 738
FT D:SULFID 740 BY SIMILARITY. 749
FT D:SULFID 756 BY SIMILARITY. 767
FT D:SULFID 761 BY SIMILARITY. 776
FT D:SULFID 778 BY SIMILARITY. 787
FT D:SULFID 794 BY SIMILARITY. 805
FT D:SULFID 799 BY SIMILARITY. 814
FT D:SULFID 816 BY SIMILARITY. 825
FT D:SULFID 832 BY SIMILARITY. 843
FT D:SULFID 837 BY SIMILARITY. 854
FT D:SULFID 856 BY SIMILARITY. 865
FT D:SULFID 872 BY SIMILARITY. 883
FT D:SULFID 877 BY SIMILARITY. 892
FT D:SULFID 894 BY SIMILARITY. 903
FT D:SULFID 910 BY SIMILARITY. 921
FT D:SULFID 915 BY SIMILARITY. 930
FT D:SULFID 932 BY SIMILARITY. 941
FT D:SULFID 986 BY SIMILARITY. 997
FT D:SULFID 991 BY SIMILARITY. 1006
FT D:SULFID 1008 BY SIMILARITY. 1017
FT D:SULFID 1024 BY SIMILARITY. 1035
FT D:SULFID 1029 BY SIMILARITY. 1044
FT D:SULFID 1046 BY SIMILARITY. 1055
FT D:SULFID 1062 BY SIMILARITY. 1073
FT D:SULFID 1067 BY SIMILARITY. 1082
FT D:SULFID 1084 BY SIMILARITY. 1093
FT D:SULFID 1100 BY SIMILARITY. 1121
FT D:SULFID 1115 BY SIMILARITY. 1130

Alignment Scores:
Pred. No.: 15.5
Score: 66.50
Percent Similarity: 39.06%
Best Local Similarity: 31.25%

Length: 2524
Matches: 20
Conservative: 5
Mismatch: 26

```

```

Query Match: 9.64% Indels: 13
DB: 1 Gaps: 3

M64347 (1-385) x NOTC_XENLA (1-2524)

QY 201 TCGTAAAGTGTAACTCTACAAACAATACATCTTGTAAACTCA-----ATA 151
    ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 176 CysLysGlnAspIleAsnGluCysSerGlnAsnProCysLysAsnGlyGlnCysile 195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 150 AATTATATATATATATATATAAATTGTAACATCTTAATAACATCGGAACCTGCACACA 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 AsnGluPheGlySerTyrArgCysThrCys-----GlnAsnArgPheThr 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 90 GGG-----CCGGCCCTCCCTGGAAACCGTCTCTCTGCTGGACACACA 46
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 GlyArgAsnCysAspGluProTyrValProCysAsnProSerProCysLeuAsnGlygly 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 45 GCAATTAGAAGA 34
Db 231 ThrCysArgGln 234

RESULT 14
AAAL HUMAN STANDARD, PRT; 523 AA.
ID AAAL HUMAN STANDARD, PRT; 523 AA.
AC Q9NS82.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Asc-type amino acid transporter 1 (Asc-1).
GN SLC7A10 OR ASC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RP TISSUE=Brain;
RX MEDLINE=20323545; PubMed=10863037;
RA Nakauchi J., Matsuo H., Kim D.K., Goto A., Chairoungdua A., Cha S.H.,
RA Inatomi J., Shikawa Y., Yanaguchi K., Saito I., Endou H., Kanai Y.;
RT "Cloning and characterization of a human brain Na+-independent
RT transporter for small neutral amino acids that transports D-serine
RT with high affinity."
RL Neurosci. Lett. 287:231-235(2000).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT CSNU3 ASP-112.
RX MEDLINE=21400982; PubMed=11509015;
RA Leclerc D., Wu Q., Ellis J.R., Goodyer P., Rozen R.;
RT "Is the SLC7A10 gene on chromosome 19 a candidate locus for
RT cystinuria?"
RL Mol. Genet. Metab. 73:333-339(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Bassi M.T., Borsani G., Nunes V., Palacin M.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch).

CC	EMBL; AC006836; AAD20081.1; --
DR	PIR; D84452; D84452.
DR	InterPro; IPR007508; DUF516.
DR	Pfam; PF04414; DUF516; 1.
DR	Hypothetical protein.
DR	SEQUENCE 361 AA; 39774 MW; A8D433ADB7F4AD06 CRC64;

Alignment Scores:		
Pred. No.:	19	Length: 361
Score:	65.00	Matches: 10
Percent Similarity:	54.17%	Conservative: 3
Best Local Similarity:	41.67%	Mismatches: 11
Query Match:	9.26%	Indels: 0
DB:	1.	Gaps: 0

M64347 (1-385) x U204 ARATH (1-361)

[illegible]

Search completed: September 22, 2004, 10:51:23
Job time : 23.5 secs

This Page Blank (uspto)


```

DR EMBL; AB008036; AAK6874.1; -.
DR PIR; A97490; A97490.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0016988; P:cell wall catabolism; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR002482; LysM.
DR SMART; SM005829; Sug_transporter.
DR PROSITE; PS00257; LysM; 1.
DR PROSITE; PS00216; SUGAR TRANSPORT 1; 1.
SQ SEQUENCE 674 AA; 69457 MW; 5134DAA4818141BC CRC64;

Alignment Scores:
Pred. No.: 0.295 Length: 674
Score: 85.50 Matches: 27
Percent Similarity: 43.59% Conservative: 7
Best Local Similarity: 34.62% Mismatches: 25
Query Match: 12.39% Indels: 19
DB: 16 Gaps: 3

M64347 (1-385) x Q8USE6 (1-674)
QY 376 CCACATGCTGGGCCCCAAGAA-----GAG 350
Db 310 ProProAlaLeuGlnProGlnAlaThrProSerGlnProSerAlaAlaThrSerGlu 329
QY 349 ACCACCTGAGCCATGGCTGAGGCAAGCAAGGACGCTGCCAGACTCAG----- 296
Db 330 ThrThraAlaAsnThrProValProSerSerGlyThrAlaAlaProThrAlaAlaGly 349
QY 295 -----GGCCAGTAACA-----GTACAGAACGACCACTGAATTCACGGCT 254
Db 350 SerThrAlaValAlaProLeuThrSerSerValGlnValThrAlaValGluPheGluGly 369
QY 253 TCCCTCCCAAGCTTTGAAGGTAGCAGTCCAGGCTATAAACTCTAGAACGATTG 200
Db 370 SerIysIlePheValAlaGlySerAlaProGlyGlySerThrValArgAlaLeu 387

RESULT 2
Q96T16 PRELIMINARY; PRT; 504 AA.
AC Q96T16;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ14504.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
KW EMBL; AK027410; BAB55094.1; -.
KW Hypothetical protein.
SQ SEQUENCE 504 AA; 56917 MW; 72B57E6A7F900698 CRC64;

Alignment Scores:
Pred. No.: 0.577 Length: 504
Score: 83.00 Matches: 26
Percent Similarity: 47.14% Conservative: 7
Best Local Similarity: 37.14% Mismatches: 19
Query Match: 12.03% Indels: 18
DB: 4 Gaps: 3

```

```

M64347 (1-385) x Q96T16 (1-504)
QY 364 GGGCCCCCAGAGAGACACCCCTGAGCATGGCTCGAGGCAAGCAGGACGACCTGCC 305
Db 425 GlyProGlnGluArgThrProCys-----GlyProAlaLeuArgGluGlyGluAla 442
QY 304 CAGACTCAGGCGCCAGTAACAGTACAGAAC-----GAACCA 269
Db 443 AlaValGluGlyProCysProSerGlnGluSerGlnGluGluAsnProGluPro 462
QY 268 ACTGAATTCACGGCTTC-----CTCCAAGCTTTGAAGGT 233
Db 463 ThrGluAspGluArgSerGluGlyValGluValGluValGluSerCysGlnGly 482
QY 232 AGCAGTCACGAGCTATATAAACTCTAGAACCA 203
Db 483 SerSerAsnGlyAlaGlnAspGlnGluAla 492

RESULT 3
Q8IV67 PRELIMINARY; PRT; 440 AA.
AC Q8IV67;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Solute carrier family 13 (Sodium-dependent dicarboxylate transporter),
DE member 2.
GN SLC13A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hagos Y., Bahn A., Burchhardt G.;
RT "Molecular cloning and characterization of human sodium dicarboxylate
RT cotransporter 2 (hNDC2)".
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ489980; CAD34590.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006814; P:sodium ion transport; IEA.
DR InterPro; IPR001898; Na/sul symtp.
DR Pfam; PF00939; Na_sulph_symp; 1.
SQ SEQUENCE 440 AA; 47630 MW; DF0CDD54D9D1CC2 CRC64;

Alignment Scores:
Pred. No.: 1.14 Length: 440
Score: 80.50 Matches: 23
Percent Similarity: 47.83% Conservative: 10
Best Local Similarity: 33.33% Mismatches: 27
Query Match: 11.67% Indels: 9
DB: 4 Gaps: 2

M64347 (1-385) x Q8IV67 (1-440)
QY 385 CACCTCTGGCCACCATGCTGCGGCCCAAGAGAGACACCCCTGAGCCTGCCCTGCA 326
Db 318 HisLeu-----CysGlyHisProAlaGluAsnSerIleLeuSerProSerProAla 334
QY 325 GCGAAGCAAGGACAGCTGCCAGTACAGGCGCCAGTACAGTACAGAACGACCAACT 266
Db 335 GlyLeuGluGlyAsnProGlyGluSerAlaGlyAlaValArgValAspGlyGlu----- 352
QY 265 GAATTCACGGCTTCCTCCCAAGCTTTGAAGGTAGTCCAGCTATAAACTCTAGAA 206
Db 353 -----AlaHisGlyAlaLeuAlaArgSerAlaProGlySerHisProLeuAsp 368
QY 205 GCATTGGCTAAGAGAGTGTAAAGTCTAC 179
Db 369 PheValLeuAlaArgCysArgValHis 377

```

GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR006820; Caudal_act.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR000047; HTH_lambpressor.

DR Pfam; PF04731; Caudal_act; 1.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR PRINTS; PR00031; HTHREPRESSR.

DR PRODOM; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.

DR PROSITE; PS00071; HOMEBOX 2; 1.

KW DNA-binding; Homeobox; Nuclear protein.

SQ SEQUENCE 263 AA; 28817 MW; E87B71EBA1A7A695 CRC64;

Alignment Scores:

Pred. No.:	3.32	Length:	263
Score:	76.50	Matches:	38
Percent Similarity:	37.50%	Conservative:	13
Best Local Similarity:	27.94%	Mismatches:	53
Query Match:	11.09%	Indels:	32
DB:	13	Gaps:	6

M64347 (1-385) x Q90254 (1-263)

QY 379 TGGCCACCATGCACTGGGCCCCAAGAGAGACACCCCTGACGCTGCTGCA---GGC 323

DB 58 TrpGlyProHisTyrGlyProGlnArgGluAspTrpAsnAlaTyrGlyProCysProGly 77

QY 322 AAGCAAGGACAGCTGCCAGACTCAG-----GGCCAGTAACAGTACAGAAC 275

DB 78 ProSerGlyThrAlaAlaAlaAlaGlnLeuSerGlySerSerProGlyGlnGlySerTyr 97

QY 274 GAACCAACTGATTCACGGCTTCCTCCCAAGCTTTGAAAGGTAGCAGTCCA----- 224

DB 98 SerProAlaGluTyrSerSerLeuHisProAlaAlaProGlyGlyLeuProProAlaGly 117

QY 223 -----GGCTATAAACTCTAGACCATG 200

DB 118 ThrValGlyThrGlnGlnValSerProThrAsnGlnArgHisSerSerTyrGluTrpMet 137

QY 199 CGTAGAGAGTGTAAAGTCTACACAAATACATCTGTATAAACTCAATAAATATATATA 140

DB 138 ArgLys-ThrVal-----ProThrAsnThrThrGlyLysThrArgThrLysGluLysTyr 155

QY 139 TAGATATATATAAACTGTAAACATCTATAACATCGGAACCTGCACACAGGCGCGCC 80

DB 155 rArgValValTyrThrAsp-----HisGlnArgLeuGluLeuGlu----- 168

QY 79 TCCCTGGAAACGCTCTCCCTGCTGGGACACACAGCANTTAGACA 34

DB 169 -----LysGluPheHisCysAsnArgTyrIleThrIleArg 181

RESULT 6

Q17806 PRELIMINARY; PRT; 263 AA.

ID Q17806

AC Q17806; PRELIMINARY; PRT; 263 AA.

DT 01-NOV-1996 (TREMELrel. 01, Created)

DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE C08A9.3 protein.

GN C08A9.3

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=94150118; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

DR InterPro; IPR006820; Caudal_act.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR000047; HTH_lambpressor.

DR Pfam; PF04731; Caudal_act; 1.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR PRINTS; PR00031; HTHREPRESSR.

DR PRODOM; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX 1; 1.

DR PROSITE; PS00071; HOMEBOX 2; 1.

KW DNA-binding; Homeobox; Nuclear protein.

SQ SEQUENCE 263 AA; 28817 MW; E87B71EBA1A7A695 CRC64;

Alignment Scores:

Pred. No.:	1.34	Length:	562
Score:	80.00 <td>Matches:</td> <td>26</td>	Matches:	26
Percent Similarity:	45.71% <td>Conservative:</td> <td>6</td>	Conservative:	6
Best Local Similarity:	37.14% <td>Mismatches:</td> <td>20</td>	Mismatches:	20
Query Match:	11.59% <td>Indels:</td> <td>18</td>	Indels:	18
DB:	4	Gaps:	3

M64347 (1-385) x Q86W50 (1-562)

QY 364 GGGCCCCAGAGAGACACCCCTGAGCCATGGCCCTGAGGCAAGCAAGGACAGCTGCC 305

DB 425 GlyProGlnGluArgThrProCys-----GlyProAlaLeuArgGluGlyGluAla 442

QY 304 CAGACTCAGGCGCCAGTACAGTACAGAAC-----GAACCA 269

DB 443 AlaValGluGlyProCysProSerGlnGluSerLeuSerGlnGluGluAsnProGluPro 462

QY 268 ACTGAATTCAGCGCTTCC-----CTCAAGCTTTGAAGCT 233

DB 463 ThrGluAspGluArgSerGluGluLysGlyGlyValGluValLeuGluAsnCysGlnGly 482

QY 232 AGCAGTCCAGGCTATAAACTCTAGAACGA 203

DB 483 SerSerAsnGlyAlaGlnAspGlnGluAla 492

RESULT 5

Q90254 PRELIMINARY; PRT; 263 AA.

ID Q90254

AC Q90254; PRELIMINARY; PRT; 263 AA.

DT 01-DEC-2001 (TREMELrel. 19, Created)

DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Caudal-type homeobox protein CDXB.

GN CDXB.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21391686; PubMed=11500978;

RA Ehrman L.A., Yutzy K.E.;

RT "Anterior expression of the caudal homologue cCdx-B activates a posterior genetic program in avian embryos.";

RL Dev. Dyn. 221:412-421(2001).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

DR EMBL; AF353624; AA338602.1; -.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003700; P:transcription factor activity; IEA.

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMuray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U42844; AB53817.1; --
 DR PIR; TJ5453; TL5453
 DR WORMPEP; C08A9.3; C506785.
 SQ SEQUENCE 263 AA; 30460 MW; EDF6EA44080AD135 CRC64;

Alignment Scores:
 Pred. No.: 5.78 Length: 263
 Score: 74.50 Matches: 29
 Percent Similarity: 39.29% Conservative: 15
 Best Local Similarity: 25.89% Mismatches: 25
 Query Match: 10.61% Indels: 43
 DB: 5 Gaps: 5

M64347 (1-385) x Q17806 (1-263)
 QY 28 ACAATCTTCTATTGCTGCTGCTGCCAGGAGAGAGCGTTCCAGGAGGCGCGG 87
 Db 140 SerAsnSerAlaIleCysLeu-----GlyArgGlyGln 151
 QY 88 CCCTGTGTGCAGGTTCCGATGTTATTAGATGTTACAAAGTTTATATATATATATA 147
 Db 152 -----IleGlyIleTyrLeu 156
 QY 148 ATTATTGAGTTTTACAGATGT----- 171
 Db 157 ValTyrTyrAlaValGlnLysCysArgPheGluArgGlnSerPheThrLeuPheTyrLys 176
 QY 172 ATTTGTTGTAGACATAACACTCTTACGCAATGCTTCTAGAGTTTATACCTGGACTGC 231
 Db 177 IleCysCysThrLeuIlePheIleValPheMetLeuMetGlu-----CysLeuAsnArg 194
 QY 232 TACCTTTCAAAGCTTGGAGGAGCGCTGAATTCAGTTGTTGCTTCTGTACTGTACTG 291
 Db 195 TyrLeuAlaAsn-----PheMetLeuThrTyrAsnValLeuLeuThr 208
 QY 292 GGCCCTCAGTCTGGCAGCTGCTCCCTGCTGCTGCTG 327
 Db 209 ProAlaLysSerArgGlnLeuAsnMetAspCysVal 220

RESULT 7

Q96727 ID Q96727 PRELIMINARY; PRT; 867 AA.
 AC Q96727
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN BBL3.060.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Algen V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AU513462; CAD11405.1; --
 DR InterPro; IPR000433; Znf_ZZ.
 DR Pfam; PF00569; ZF_ZZ_2.
 DR SMART; SM00291; Znf_ZZ_4.
 DR PROSITE; PS0135; ZF_ZZ_2; 3.
 KW Hypothetical protein_
 SQ SEQUENCE 867 AA; 95865 MW; CE267925A35FE211 CRC64;

Alignment Scores:
 Pred. No.: 9.78 Length: 867
 Score: 73.00 Matches: 20
 Percent Similarity: 53.23% Conservative: 13
 Best Local Similarity: 32.26% Mismatches: 21
 Query Match: 10.58% Indels: 8
 DB: 3 Gaps: 2

M64347 (1-385) x Q96727 (1-867)
 QY 361 CCCCAAGAGAGAGACCACTGAGCATGGCTGAGCAAGGAGGAGGAGGAGGAG 302
 Db 110 ProGlnAspThrThrGluAlaLys---ProAlaSerValGluGlyGluAlaThrGlu 128
 QY 301 ACTCAGGCGCCAGTAAACA-----GTACAGAACGAAACCAACTGAA 263
 Db 129 ThrGluGlyAspAlaSerProSerAsnProThrSerProValTyrAspGluThrGlu 148
 QY 262 TTCAGCGCTTCCCTCCCAAGCTTTGAAAGGTAGCTCCAGGCTATAAAGCTAGAGCA 203
 Db 149 ValThrLysSerIleGluThrAspAlaSerProSerLysThrPheGluSerLeuGlnAla 168
 QY 202 TTGCGT 197
 Db 169 GlnArg 170

RESULT 8

Q9FQ80 ID Q9FQ80 PRELIMINARY; PRT; 955 AA.
 AC Q9FQ80
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase).
 GN PPC.
 OS Chloris gayana.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Chloridoideae; Chlorideae; Chloris.
 OX NCBI_TaxID=110876;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fully expanded leaves;
 RA Blaesing O.E., Gowik U., Wolf B., Westhoff P.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
 CC -!- SOURCE FOR THE TRICARBOXYLIC ACID CYCLE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: PHOSPHATE + OXALOACETATE = H(2)O +
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SIMILARITY: BELONGS TO THE PEPCase FAMILY.
 DR EMBL; AF268091; AAG42288.1; --
 DR HSP; P00864; 1PIY.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008964; P:phosphoenolpyruvate carboxylase activity; IEA.
 DR GO; GO:0015977; P:carbon utilization by fixation of carbon di...; IEA.

```

M64347 (1-385) x Q19613 (1-474)

Qy      11 CAAGCTGGTATTTCATACAAATCTTCTAATTGCTGTG-----TGTCCCAGGCAG 61
       |||::|||::: ||| |||:::|||||:::
Db     341 GLSserGlyMetIleileSerPheLeulleLeualailePheleutrPilleAspTyHis 360
Qy      62 GGAGACGGTTTCAGGAGGCGGCCCTGTGTGCAGGTTCGCATGTATTAGATGTA 121
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     361 GlyThrGlyPhe-----Leu 365
Qy     122 CAAGTTTATATATATCATATATATAATATTTATGAGTTTTTACAAGATGATTTGTTGTA 181
       |||::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     366 IleValFyrLeupheGlyThrMet---PheIlegluFyrThrTrpAspAla----- 381
Qy     182 GACTTACACTTCTTACGCAATGCTTCTAGAGTTTATAGCCTGGACTGCTACCTTTCAA 241
       ||| ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db     382 -----ValFyrLeuCysAlaIleGlu 388
Qy     242 AGCTTGGAGGAAGCCGTGAATTCAGTTGGTTCGTTCTGTACTGTTACTGGGCCCTGAGT 301
       |||::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     389 SerMetgluThrSerSerArg-----AlaSerAla 398
Qy     302 CTGGGACGTGTCCCTTCCTGCTGCTGCAGGGCCAGTCTCAGGTGGTCTCTCTCTTGGGG 361
       |||::|||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     399 ValglySerCysSerLeu-----MetaalaarglleGlySerLeuleuAla 413
Qy     362 CCC-----AGTGCATGTGGCCA 379
       ||| ::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db     414 ProPheleuthrTyAlaAsnThrTrpPro 424

RESULT 10
Q97UC7 PRELIMINARY; PRT; 324 AA.
AC Q97UC7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SSO3088.
GN SSO3088.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong L., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland J.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE068899; AAAK3194.1; --
DR PIR; C90492; C90492.
DR InterPro; IPR001455; UPF0033.
DR PROSITE; PS01148; UPF0033; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 324 AA; 38040 MW; 861EF3865FDB448E CRC64;

Alignment Scores:
Pred. No.: 15.6 Length: 324
Score: 71.00 Matches: 26
Percent Similarity: 41.30% Conservative: 12
Best Local Similarity: 28.28% Mismatches: 18
Query Match: 10.11% Indels: 36
          17 Gaps: 6
          08.

```


OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Shan Y.X., Yu L.;
RT	"Cloning, characterization and location of a novel human gene
RL	containing an EGF domain.";
RE	Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL/AZ80362; AAP5084.1; --
SQ	SEQUENCE 2386 AA; 254570 MW; 691CE4739C9D3EBB CRC64;

Alignment Scores:	
Pred. No.:	18.7
Score:	Length: 2386
Percent Similarity:	71.00
Best Local Similarity:	Conservative: 25
Query Match:	Mismatches: 25
DB:	Indels: 16
	Gaps: 3

M64347	(1-385) x Q7Z7M0 (1-2386)
Qy	182 GACTTAAACACITCTTTACGCAATGCTTCTAGAGTTTTATAGCTGGACTGCTACCTTTCAA 241
Dd	360 GluLeuThrLeuLeu-----TpaspargThr----- 368
Qy	242 AGCTTGAGGGAAGCCCGAATCAGTGTGTTCTGTCGTACTGTATGGGCC----- 295
Dd	369 GlyValProGlyGlySerGluIleSerPhePheLeuGluProTyArgSerSer 388
Qy	296 ---CTGAGTCTGGGCAGCTGTCCCCTGCTTGCTGCAGGGCCCATGCTCAGGGTGCTCTC 352
Dd	389 CysThrSerTyRSerSerCys--LeuGlyCysLeuAlaaspGlnGlyCysGlyTrpCysL 408
Qy	353 TTCTTTGGGGCCCAAGTCATGTGTGCCAGAGT 384
Dd	408 euthr-SerAlaThrCysHisLeuArgGlnGly 418

```

RESULT 14
ID Q9PKJ7 PRELIMINARY; PRT; 360 AA.
AC Q9PKJ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT DT
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein TC0468.
DC TC0468.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83560;
[1]
RN SEQUENCE FROM N.A.
RP
RC STRAIN=McpN / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gall S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT RT pneumoniae A839";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AF002315; AAF39317.1; -.
DR FIC; C81699; C81699.
DR TIGR; TC0468; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 360 AA; 40411 MW; 7E0621028581F5FC CRC64;

Alignment Scores:
Pred. No.: 18.1 Length: 360
Score: 70.50 Matches: 32
Percent Similarity: 24.93% Conservative: 19
Best Local Similarity: 21.92% Mismatches: 28
Query Match: 10.04% Indels: 67
DB: 16 Gaps: 7

```

